OPTIMAL INTEGRATION OF SOLAR HOME SYSTEMS AND APPLIANCE SCHEDULING FOR RESIDENTIAL HOMES UNDER SEVERE NATIONAL LOAD SHEDDING

by

Sakhile Nqobile Twala

Submitted in partial fulfillment of the requirements for the degree
Master of Engineering (Electrical Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

September 2024

© University of Pretoria

SUMMARY

OPTIMAL INTEGRATION OF SOLAR HOME SYSTEMS AND APPLIANCE SCHEDULING FOR RESIDENTIAL HOMES UNDER SEVERE NATIONAL LOAD SHEDDING

by

Sakhile Nqobile Twala

Supervisor:	Prof. X. Ye
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Master of Engineering (Electrical Engineering)
Keywords:	Load shedding, Inconvenience, Optimal scheduling and sizing strategies, K-Nearest Neighbour (KNN), Multi-objective mixed integer nonlinear optimisation (MOMINLP), Load profile clustering, K-means clustering

In developing countries such as South Africa, users experienced more than 1 030 hours of load shedding outages in just the first half of 2023 due to inadequate power supply from the national grid. Residential homes that cannot afford to take action to mitigate the challenges of load shedding are severely inconvenienced as they have to reschedule their demand involuntarily. This study presents optimal strategies to guide households in determining suitable scheduling and sizing solutions for solar home systems to mitigate the inconvenience experienced by residents due to load shedding. To start with, this study predicts the load shedding stages that are used as input for optimal strategies using the K-Nearest Neighbour (KNN) algorithm. Based on an accurate forecast of future load shedding outages, this study formulates inconvenience for residents and loss of power supply during load shedding as the objective function. When solving the multi-objective optimisation problem, four different strategies to fight against load shedding are identified, namely (1) optimal home appliance scheduling (HAS) under load shedding; (2) optimal HAS supported by solar panels; (3) optimal HAS supported by batteries, and (4) optimal HAS supported by the solar home system (SHS) with both solar panels and

© University of Pretoria

batteries. Among these strategies, appliance scheduling with an optimally sized 9.6 kWh battery and a 2.74 kWp panel array of five 550 Wp panels, eliminates the loss of power supply probability and reduces inconvenience by 92% when tested in the South African load shedding cases in 2023.

More than 18.5 million households in South Africa are affected by load shedding. This results in a potential of 18.5 million unique load profiles. Creating unique optimal solar home systems for each household without evaluating similarities in their load profiles risks duplicating solar home system strategies. Clustering is an unsupervised machine learning method that can group households based on their inherent similarities, minimising intracluster similarity and maximising intercluster dissimilarity.

K-means clustering is used in a case study of 781 South African households metered for a year, forming representative clusters of energy demand profiles to identify optimal strategies for multiple households that minimise the impact of load shedding. Three load clusters are identified as optimal, using the Davies-Bouldin criterion to minimise the ratio of within- and between- cluster distances. From K-means clustering, 43% of households are clustered in a low-energy demand load profile with an average daily energy consumption of 4.5 kWh, 42% in a medium energy and 15% in a high-energy demand profile, with an average daily energy usage of 10.8 kWh and 22.1 kWh, respectively. Additionally, based on a 94.4% accurate, hourly, one-year-ahead prediction of load shedding outages using the KNN algorithm, we formulate each cluster's inconvenience and loss of power supply due to load shedding as a multiobjective mixed-integer nonlinear optimisation problem. The results show that optimal scheduling of a low, medium and high energy consumption cluster with optimally sized 2.4 kWh and 0.39 kWp, 4.8 kWh and 0.78 kWp and 7.2 kWh and 1.17 kWp, battery and panel arrays, respectively, minimises the loss of power supply and substantially reduces the inconvenience of involuntary rescheduling by 86.8%, 71.1% and 86.4% for clusters 1, 2, and 3, respectively.

© University of Pretoria

LIST OF ABBREVIATIONS

AA	Adaptable appliances
ANN	Artificial neural network
BESS	Battery energy storage system
CA	Continuous appliances
CN	Correct negative classifications
COA	Coordinated appliances
CP	Correct positive classifications
CSP	Concentrated solar power
Eskom	South African Electricity Supply Commission
FDR	False discovery rates
HAS	Home appliance scheduling
IA	Interruptible appliances
ILS	Interruptible load shed
IOS	Interruption of supply
IPP	Independent power producer
KNN	K-Nearest Neighbour
LPS	Loss of power supply
LPSP	Loss of power supply probability
LS	Load shedding
MINLP	Mixed integer nonlinear programming
MLR	Manual load reduction
MO	Multi objective programming
MRMR	Minimum redundancy maximum relevance
OCGT	Open cycle gas turbine
OCLF	Other capacity loss factor of the utility's plants
PCLF	Planned capacity loss factor of the utility's plants
PPV	Positive predictive values
PV	Photovoltaic array
RSA	Republic of South Africa
SCO	Synchronous condenser operation
SOE	State of energy
SVM	Support vector machine

© University of Pretoria

UCLF	Unplanned capacity loss factor of the utility's plants
WN	Wrong negative classifications
WP	Wrong positive classifications

© University of Pretoria

LIST OF SYMBOLS

Δt	Appliance scheduling interval
η	The charging and discharging efficiency
A_c	Cross sectional area of PV as a sizing variable
c	K-means cluster centroids
C	Integer class labels for load shedding stages
C_b	Battery capacity as a sizing variable
d	Manhattan distance metric
D_j	The jth appliance operational duration
e	Energy consumption of residential households
E_j	The jth appliance operational ending time
j	Appliance index
K	Number of nearest neighbours for KNN
k	Time index
N	Total sampling interval
P_b	Battery power as a continuous variable (kW)
P_{gr}	Grid power as a continuous variable (kW)
P_L	Load demand as a continuous variable (kW)
P_{PVA}	Maximum PV power generated for A_c (kW)
P_{PV}	PV power as a continuous variable (kW)
s	The silhouette value ranging from -1 to 1
S_j	The jth appliance operational starting time
u_j^{bl}	Baseline binary switch of appliance j
u_j^{opt}	Optimised binary switch of appliance j
u_j	The jth appliance switch as a binary variable
u_{LS}	Load shedding outages as a binary variable
w_1	Loss of power supply weighting factor
w_2	Inconvenience weighting factor
x_{LS}	Load shedding stages as an integer variable
y	Nearest neighbours for KNN

© University of Pretoria

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION **1**
1.1 PROBLEM STATEMENT 1
1.1.1 Context of the problem 1
1.1.2 Research gap 2
1.2 RESEARCH QUESTIONS 3
1.3 APPROACH 3
1.4 RESEARCH GOALS 3
1.5 RESEARCH CONTRIBUTION 4
1.6 RESEARCH OUTPUTS 4
1.7 OVERVIEW OF STUDY 4

CHAPTER 2 LITERATURE STUDY AND ANALYSIS **6**
2.1 LITERATURE STUDY ON SOLAR HOME SYSTEMS AND THEIR STRATEGIES IN RESPONSE TO LOAD SHEDDING 6
2.2 ANALYSIS OF SEVERE NATIONAL LOAD SHEDDING 9

CHAPTER 3 PROBLEM STATEMENT AND FORMULATION **12**
3.1 LOAD SHEDDING MODEL 14
3.2 PV SYSTEM MODEL 24
3.3 BATTERY ENERGY STORAGE SYSTEM MODEL 25
3.4 LOSS OF POWER SUPPLY 25
3.5 INCONVENIENCE BY LOAD SHEDDING 26
3.6 HOME APPLIANCES SCHEDULING MODEL 26
3.7 DECISION VARIABLES 26
3.8 CONSTRAINTS 27
3.9 OPTIMISATION ALGORITHM 28

© University of Pretoria

CHAPTER 4 CASE STUDY: EVALUATING STRATEGIES TO MINIMIZE THE IMPACT OF LOAD SHEDDING ON A WORKING-CLASS HOUSEHOLD 30

4.1 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING 33

4.2 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING AND A PHOTOVOLTAIC SOLAR PANEL ARRAY 34

4.3 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING AND BATTERY ENERGY STORAGE 36

4.4 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING, A SOLAR PANEL ARRAY AND BATTERY ENERGY STORAGE 39

CHAPTER 5 LOAD PROFILE CLUSTERING 42

5.1 PROPOSED METHOD 44

5.2 K-MEANS CLUSTERING DISTANCE SELECTION 48

5.3 K-MEANS CLUSTERING OF LOAD PROFILES CONSIDERING LOAD SHEDDING 55

5.4 OPTIMISATION FORMULATION 56

CHAPTER 6 CASE STUDY: CLUSTERING RESIDENTIAL HOUSEHOLD LOAD PROFILES 60

6.1 OVERVIEW OF THE METERED HOUSEHOLDS 60

6.2 DISAGGREGATING APPLIANCE USAGE FROM METERED HOUSEHOLDS . 61

6.3 LOAD SHEDDING IN HATFIELD 64

6.4 OPTIMAL NUMBER OF CLUSTERS K, AND OPTIMAL LOAD PROFILE CLUSTERING 66

6.5 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR LOW ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING 74

6.6 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR MEDIUM ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING 78

6.7 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR HIGH ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING 81

6.8 MINIMISING INVOLUNTARY INCONVENIENCE FOR CLUSTERED RESIDENTIAL HOUSEHOLDS UNDER LOAD SHEDDING 84

CHAPTER 7 CONCLUSION 91

REFERENCES 94

© University of Pretoria

LIST OF FIGURES

2.1 The annual load shedding hours in Tshwane, South Africa. 10
2.2 The monthly load shedding hours in Tshwane, South Africa. 11

3.1 Illustrative schematic for home appliances control. 13
3.2 A scatter plot of dispatchable generation and load shedding, versus residual demand. 15
3.3 A box and whisker plot on load shedding stages and manual load reduction (load shedding outages) on Eskom's dataset. 16
3.4 A heat map of Pearson's correlation coefficients on Eskom's dataset. 17
3.5 Feature importance scores when predicting load shedding stages sorted using a Minimum Reduncancy Maximum Relevance algorithm in MATLAB on Eskom's dataset. 18
3.6 The confusion chart of Mahalanobis KNN predictions of the load shedding stages in Hatfield, South Africa. 20
3.7 Shapley values to interpret a single load shedding prediction by the model for the KNN features of year 2023, month 8, day 29 and hour 15, with the KNN output predicting stage 4 which is the true load shedding stage. 23
3.8 The average load shedding hours per day in Hatfield, South Africa. 24

4.1 The load shedding outage profile based on the average load shedding outage duration and on the hourly load shedding probability predictions for 2023. 31
4.2 (a) The optimal power demand for scheduling with load shedding; (b) with the optimal power flow management of the energy sources; (c) and the inconvenience. 34
4.3 (a) The optimal power demand for scheduling with load shedding using Figure 4.1, and a photovoltaic solar panel array (b) with the optimal power flow management of the energy sources (c) and the inconvenience. 35

© University of Pretoria

4.4 (a) The optimal power demand for scheduling with load shedding using Figure 4.1, and battery energy storage, (b) with the optimal power flow management of the energy sources (c) and the inconvenience. 37
4.5 (a) The Pareto set of solutions for sizing the batteries when scheduling with load shedding and battery energy storage, and (b) the Pareto set of solutions for sizing the batteries and panels when scheduling with load shedding, panels and batteries using Figure 4.1. 38
4.6 The feasible criterion region to determine the sizing decision variables subject to the constraints of scheduling with load shedding using Figure 4.1. 39
4.7 (a) The optimal power demand for scheduling with load shedding using Figure 4.1, a solar panel array and battery energy storage, (b) with the optimal power flow management of the energy sources (c) and the inconvenience. 41

5.1 Illustration schematic for proposed residential power flow management with renewable energy integration considering load shedding in clustered residential homes. 46
5.2 An illustration of the proposed household energy consumption clustering using K-means. 59

6.1 The annual load shedding hours and average electricity price in Hatfield, South Africa. 65
6.2 Determining the optimal number of clusters K, in K-means clustering, based on the Davies-Bouldin criterion. 67
6.3 The silhouette values assessing the degree of separation among three clusters using the cityblock distance metric for K-means clustering. 68
6.4 The box-and-whisker plot of K-means clustered households shows load profiles classified into low, medium, and high energy consumption clusters 1, 2 and 3, respectively. 71
6.5 A histogram of average daily energy consumption in K-means clusters, divided into 20 bins, with some bins overlapping with other clusters. Blue is for low-energy households, red for medium-energy-consuming households, and green for high-energy households. 73
6.6 The average daily load profiles for cluster 1 households using K-means clustering. . . 75
6.7 (a) The optimal scheduling of cluster 1's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling. 77
6.8 The average daily load profiles for cluster 2 households using K-means clustering. . . 78

© University of Pretoria

6.9 (a) The optimal scheduling of cluster 2's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling. 80
6.10 The average daily load profiles for cluster 3 households using K-means clustering. . . 82
6.11 (a) The optimal scheduling of cluster 3's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling. 83

© University of Pretoria

LIST OF TABLES

3.1 The accuracy of different prediction methods when compared to the KNN, to forecast the load shedding stages in 2023, Hatfield, South Africa. 22
3.2 SCIP settings. 29

4.1 Specifications of the hybrid power system. 31
4.2 Appliance specifications and baseline energy usage scheduling without load shedding. 32
4.3 Weekly effects of South Africa's load shedding stages when no action is taken. . . . 32
4.4 The daily loss of power supply, inconvenience, sizing and energy management for all four strategies. 33

6.1 The number of households in each of the K-means clusters. 72
6.2 Monthly effects of South Africa's load shedding stages on low energy consuming households without solar panels or batteries. 76
6.3 Monthly effects of South Africa's load shedding stages on medium energy consuming households without solar panels or batteries. 79
6.4 Monthly effects of South Africa's load shedding stages on high energy consuming households without solar panels or batteries. 84

© University of Pretoria

CHAPTER 1 INTRODUCTION

1.1 PROBLEM STATEMENT

1.1.1 Context of the problem

In October 2020, a storm left more than 300 000 people without power in Oklahoma, in the United States [1] and in February 2021, a series of winter storms in Texas resulted in sweeping power outages that lasted for more than two weeks [2]. China experienced its worst power outages in almost a decade in the first half of 2021 and more than half of China's provinces were subjected to rationing of power at the industrial and commercial level [3]. Although the causes of power outages vary between developed and developing countries, and despite how each country categorises them, it is important to understand their impact and equip end users with effective mitigation strategies.

Reliable electricity supply is essential for economic growth, and the persistent problem of grid-scale power outages is accelerating investment in small-scale embedded generation [4]. Load shedding is a strategy for power outages that coordinates the rationing of limited electricity on a grid-scale. The study in [5] estimates that due to load shedding, Nepal suffered an economic loss of US$11 billion in terms of its Gross Domestic Product (GDP) over the nine years from 2008 to 2016. This amount is equal to Nepal's GDP in 2008. Power outages cause significant inconvenience for residential consumers. During load shedding in Zambia, 95.5% of households in low-income areas, 85.5% in middle-income areas and 73.1% in high-income areas resort to using charcoal for cooking [6]. In Zimbabwe, 60% of the residents have lost perishable food due to power outages and the proportion of energy cost to income has increased from 16% without load shedding to as high as 64% in low-density areas and 49% in high-density areas [7].

South Africa accounts for nearly 45% of Africa's total energy consumption, yet its utility infrastructure struggles to meet the nation's electricity demand [8]. Since 2008, Eskom, South Africa's state-owned

© University of Pretoria

electricity company, has struggled to meet the growing demand for electricity. The company uses electricity rationing measures to avoid nationwide blackouts [9]. The annual energy availability factor in South Africa has decreased from 2016 to 2023, and the country has experienced more than 7 431 cumulative hours of power outages [10]. This means that the country has endured more than 309 days in a state of darkness. South Africa is among several countries that face persistent power outages. For example, between June 2015 and March 2017, Zambia's national electricity utility, the Zambia Electricity Supply Corporation (ZESCO), implemented increased load shedding outages [6, 11]. These load shedding outages last between 6 and 10 hours a day, impacting industries, commercial zones, offices, and households alike. At the height of the power supply crisis in Ghana, electricity users endured up to 16 hours of daily outages, meaning they had electricity for only 8 hours a day [12].

Prior to severe load shedding, households used their appliances during times that were most convenient for them. Those who had a plan on when to use each appliance would create a home appliance schedule. The motivation to schedule appliances was to reduce electricity costs. However, if the schedule recommended an inconvenient time for use, households had the freedom to choose whether or not to follow it. Load shedding outages have taken away the ability of households to voluntarily schedule their appliance usage. Consumers have lost control over when they can use their appliances and are forced into involuntarily scheduling.

Load shedding has forced consumers to look for strategies against its involuntary scheduling. Residential customers are more affected by load shedding than industrial and commercial customers, because they are less prepared for the outages. Most households do not have diesel generators and have fewer resources to spend on backup power than commercial and industrial customers. This study reveals that residential customers are at a higher risk of load shedding, based on the probability of when the most outages occur in a 24-hour period. Due to load shedding, households reschedule their appliances to inconvenient periods before and after the outages. When there is load shedding, each appliance typically used during that time is rescheduled to a less convenient time slot. The difference between the usual operating times of the appliances and their rescheduled period is defined as inconvenience.

1.1.2 Research gap

Studies on appliance scheduling focus mainly on voluntary participation in demand-side management (DSM) [13, 14], without considering whether their outcomes are applicable when the DSM is involuntary. Electricity tariff incentives are voluntary, while load shedding outages are involuntary. The study

© University of Pretoria

in [15] estimates the cost of 99 days of load shedding in South Africa, between January and June 2015, to be R 13.72 billion. However, there are three limitations to their contribution. Firstly, the study lacks an empirical, data-driven approach to determine the probability of load shedding outages. Secondly, it does not offer solutions to mitigate these outages. Lastly, the potential bias arises from a low response rate and a reliance on subjective, non-generalisable survey data.

1.2 RESEARCH QUESTIONS

The four research questions in this study are as follows:

- How can we model the experienced load shedding outages?
- What strategies do residential homes have to minimise the inconvenience and loss of power supply caused by load shedding?
- What strategies do residential homes have to address involuntary scheduling, when they are forced to reschedule their appliances?
- When more than one household is considered, what scheduling, solar panel, and battery sizing strategies do low-energy consumption, medium-energy consumption, and high-energy-consumption households have to minimise the inconvenience and loss of power supply caused by load shedding?

1.3 APPROACH

Households under load shedding are forced to reschedule their appliance use to periods before and after outages, causing inconvenience and loss of power supply. The load shedding in South Africa exhibits the following key features: 1) it can last more than 2 hours per load shedding outage; 2) the load shedding schedule and duration change from one day to the next; ordinary residents have no way to predict the exact start and end time of load shedding, and they have no way to know the exact duration for each load shedding outage; and 3) the load shedding patterns vary across suburbs and townships throughout the nation. This makes it challenging to identify universal solutions for residents from different areas of the country to mitigate the negative impact of load shedding. To address the very practical and urgent challenges of load shedding, this study aims to identify several optimal solutions to help residential homes mitigate the negative impact of load shedding.

1.4 RESEARCH GOALS

The three research goals of this study are to:

© University of Pretoria

- Predict the load shedding outages in South Africa based on historical data of load shedding in different areas from 2014 to 2023;
- Determine the optimal appliance scheduling and solar panel and battery storage sizing strategy to minimise the load shedding inconvenience and loss of power supply for residential homes that:

  1. Cannot invest in any backup alternative;
  2. Can only invest in solar panels;
  3. Can only invest in batteries;
  4. Can invest in both solar panels and batteries.

- Cluster load profiles to determine the optimal strategies that integrate solar panel and battery sizing with scheduling for multiple households with different load profiles.

1.5 RESEARCH CONTRIBUTION

This study presents optimal strategies to guide households in determining suitable scheduling and sizing solutions for solar home systems to mitigate the inconvenience experienced by residents due to load shedding. To start with, this study predicts the load shedding stages that are used as input for the optimal strategies by using the K-Nearest Neighbour (KNN) algorithm. Based on an accurate forecast of the future load shedding patterns, this study formulates the residents' inconvenience and the loss of power supply probability during load shedding as the objective function. When solving the multi-objective optimisation problem, four different strategies to fight against load shedding are identified, namely (1) optimal home appliance scheduling (HAS) under load shedding; (2) optimal HAS supported by solar panels; (3) optimal HAS supported by batteries, and (4) optimal HAS supported by the solar home system with both solar panels and batteries.

1.6 RESEARCH OUTPUTS

S. Twala, X. Ye, X. Xia and L. Zhang, "Optimal integration of solar home systems and appliance scheduling for residential homes under severe national load shedding", *Journal of Automation and Intelligence*, vol. 2, no. 4, pp. 227-238, 2023, https://doi.org/10.1016/j.jai.2023.12.001.

1.7 OVERVIEW OF STUDY

The remainder of this study is structured as follows: Chapter 2 reviews the literature on load shedding, appliance scheduling, and sizing strategies. This chapter presents the literature study for this dissertation as well as a statistical analysis of the load shedding outages experienced in South Africa from 2015 to

© University of Pretoria

2023. Chapter 3 presents the problem statement with its formulation, the energy modelling, including the novel load shedding model, and the optimisation formulation. Chapter 4 presents the case study and the results of optimal strategies. Chapter 5 presents the clustering of load profiles. Chapter 6 presents the case study of clustered homes as well as the results on the integration of solar home systems for clustered homes. Finally, Chapter 7 provides the conclusion.

© University of Pretoria

CHAPTER 2 LITERATURE STUDY AND ANALYSIS

2.1 LITERATURE STUDY ON SOLAR HOME SYSTEMS AND THEIR STRATEGIES IN RESPONSE TO LOAD SHEDDING

Load shedding has become an increasingly common challenge in several countries due to factors such as inadequate electricity supply, ageing infrastructure, and increased demand. Although research in [16] and [17] has focused primarily on strategies to strengthen power distribution systems, there has been limited exploration of measures that end users can implement to enhance their own resilience against load shedding outages.

The strategies found in studies on scheduling, such as the strategy of using a feed-in and a time-of-use tariff as the incentives to participate in scheduling [18], the strategy of using time-of-use with a sensitivity analysis that highlights the significance of a customer's willingness to pay [19], and the strategy of multi-residential, voluntary power sharing for scheduling appliances [20], all rely on two assumptions: the first is that time-of-use and other electricity tariffs are the only motivation for scheduling; the second is that the decision about participating in scheduling is exclusively made by the customer. Our study does not make these assumptions; it emphasises that load shedding is different from tariff scheduling because the scheduling decisions during load shedding are beyond the voluntary control of customers.

There are studies focussing on the supply of energy to residential homes. This includes sizing renewable energy generators, deploying battery energy storage systems, and improving power electronic devices with more effective power flow management [21]. The strategy for sizing solar panels, wind turbines and batteries in response to load shedding in [22] is limited by the assumption that load shedding

© University of Pretoria

outages are always known by the end user, well ahead of time and with certainty. The oversimplification of the load shedding model is ineffective in addressing the experienced challenges facing end users to meet their demand due to unpredictable load shedding outages. The study in [20] presents a strategy to integrate an energy storage system with the scheduling of appliances for use during emergency outages. The integrated strategy depends on emergency operating conditions, which are modelled on a worst-case scenario that is not based on the historical outage patterns in the study's proposed region. In cases where the impact of load shedding extends beyond the home to edge data centres (EDCs) [23], the study's novel battery allocation strategy, which protects interactive workloads in EDCs from being interrupted and minimises the waiting time of batch workloads, also relies on the oversimplification of load shedding outage patterns.

Among various renewable energy sources such as solar, wind, hydropower and biofuels [24], solar energy harvesting by solar photovoltaic (PV) technology has experienced an exponential growth rate in recent years [25], [26]. This interest is unperturbed by the intermittency of solar irradiance results or discontinuities in solar PV panel output power. Energy storage systems (ESS) are important to ensure the continuous balance between supply and demand dynamics. Batteries are the most widely used energy storage systems to satisfy the power imbalance between photovoltaic power generation and load demand [27], [28]. Lead-acid and lithium-ion batteries are the most popular battery types and are usually connected to PV systems to store green energy and reduce grid energy consumption [29]. Generally, battery banks encounter irregular charging patterns and deep cycles, due to intermittent power availability and output from PVs and varying load demands. These operations decrease the battery life and increase the cost of battery replacement [30], [31]. To overcome this problem, hybrid energy storage (HESS) is employed in various applications, with the underlying assumption, however, that there is certainty in the power supply. The effectiveness of the proposed storage solutions requires a reliable grid that does not have load shedding. This modelling approach is not always applicable and is replaced by the data-driven load shedding modelling developed in Chapter 2.2 in our study.

In determining the sizing of a home energy system, some research explores optimal power operation on the supply side. For instance, economic dispatch strategies are considered for hybrid distributed energy generators such as PV/wind/battery/diesel generators to reduce energy costs [32]. Optimal switching control and intuitive designed control methods have also been considered [33]. The economic performance, technology performance and cost minimisation of a household PV-diesel-battery hybrid system within demand side management (DSM) are studied through analysis, designs and control

© University of Pretoria

strategies [34], [33], [35]. A similar analysis is conducted for a PV-wind-battery hybrid household system [36]. In [34] a mixed integer optimisation problem is formulated and solved, while in [35] a rule-based control strategy and a converter are developed so that the battery provides as much compensation as possible to help cope with variability in solar power generation and improve energy resiliency. In [36] a power management strategy was developed for a standalone wind-PV-battery energy system through maximum power point tracking (MPPT). In [36] a pitch angle controller was also used to manage the power flows between different energy sources and ESS in the home energy system. In [33] open loop and closed loop control is used to schedule the power flows of a small-scale grid-connected PV battery, resulting in cost savings and robust control performance in the home system. Furthermore, the impacts of demand response (DR) on cost minimisation and CO_2 emissions were investigated in [37], [38]. A residential energy hub framework was used for the power demand and scheduling strategy to minimise costs and reduce emissions in relation to time-of-use tarrifs [37].

Naturally, to achieve the best impact on energy efficiency, we should integrate energy efficiency solutions from both the demand and supply sides. Intensive research is also required to obtain a globally optimal solution in terms of minimised load demand, a properly sized renewable generation system, and prioritised power flow management. Residential homes are equipped with various electrical and thermal resources and appliances that consume significant amounts of energy, and therefore optimising their usage is an important aspect of energy management. The literature on this subject has focused on developing mathematical models and algorithms to optimise energy usage, considering various sources of uncertainty such as appliance usage patterns, energy prices, and weather conditions [39].

A dynamic load-priority-based scheduling strategy is a method of managing energy in a home system by prioritising different loads according to their energy needs [40]. This approach involves dynamically adjusting the priority of different energy-consuming devices based on current demand, available energy resources, and other factors, to optimise energy usage and reduce waste. The goal is to ensure that essential devices receive the energy they need while reducing the energy consumption and costs of less critical devices. The scheduling strategy in [40] considers time-of-use tariffs combined with the technique for order preference by similarity to the ideal solution (TOPSIS) and the entropy model. TOPSIS is a multi-criteria decision-making method that has been widely used in various fields, including energy management [41]. In the context of dynamic load-priority-based home appliance scheduling, TOPSIS has been used to evaluate and prioritise different energy-consuming devices based

© University of Pretoria

on their energy consumption, priority, and other relevant factors.

Existing home appliance scheduling studies focus on households that want to reduce their electricity costs and their energy consumption while minimising their inconvenience, in response to electricity tariffs. Appliance scheduling is voluntary, and the electricity company always satisfies the customer's load. The new problem with home appliance scheduling (HAS) is that scheduling is involuntary due to load shedding. The electricity company and the government impose load shedding in response to unpredictable generation capacity loss factors on a grid-scale, and the consumer cannot choose otherwise. The novelty of this study is to guide households in finding an optimal scheduling and sizing strategy in response to compulsory scheduling, to minimise inconvenience and loss of power supply probability due to load shedding. The current literature only addresses the problem of scheduling through the paradigm of the consumer choosing to participate in the demand side management. The novel research problem addressed in this study considers scheduling when consumers are forced to do so, through load shedding, and tries to find an optimal home appliance scheduling and sizing strategy in response to this new problem.

2.2 ANALYSIS OF SEVERE NATIONAL LOAD SHEDDING

This section analyses the load shedding outages from 2015 to 2023 in Tshwane, South Africa. The data in Figure 2.1 reveal that the affluent and less affluent regions experience similar outages. Interestingly, there was no load shedding in 2016 and 2017. A concerning trend emerged from 2018 to 2021, where load shedding worsened steadily each year. This culminated in an exponential increase in total load shedding hours from 2021 to 2022, a period that coincided with the ease of lockdown restrictions in the country. Most strikingly, the total number of power outages recorded from January to July 2023 has exceeded the total load shedding hours experienced in 2022, highlighting an alarming escalation in the frequency of these disruptions.

Figure 2.2 reveals the monthly load shedding duration. In particular, August presents less load shedding than any other month. Contrary to intuitive expectations, the summer months have more load shedding compared to the winter months.

© University of Pretoria

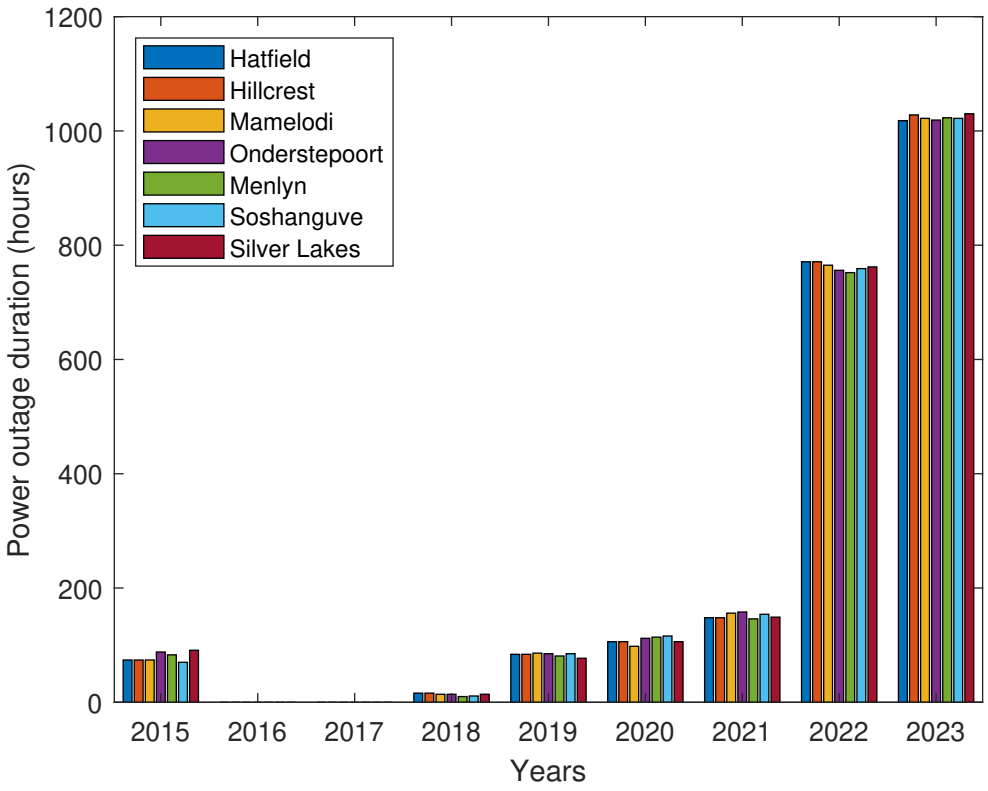

Figure 2.1. The annual load shedding hours in Tshwane, South Africa.

© University of Pretoria

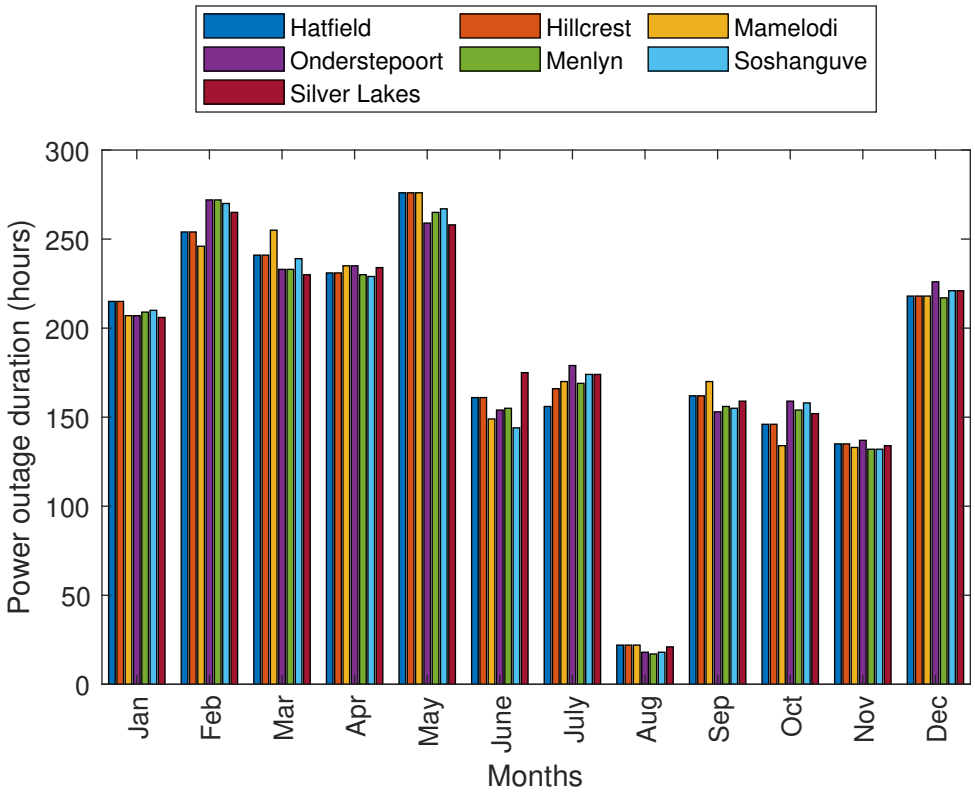

Figure 2.2. The monthly load shedding hours in Tshwane, South Africa.

© University of Pretoria

CHAPTER 3 PROBLEM STATEMENT AND FORMULATION

The load shedding in South Africa exhibits the following key features: 1) it can last more than 2 hours per load shedding outage; 2) the load shedding schedule and duration change from one day to the next; ordinary residents have no way to predict the exact start and end time of load shedding, and they have no way to know the exact duration for each load shedding outage; and 3) the load shedding patterns vary across suburbs and townships throughout the nation. This makes it challenging to identify universal solutions for residents from different areas of the country to mitigate the negative impact of load shedding.

To address the very practical and urgent challenges of load shedding, this study aims to identify several optimal solutions to help residential homes mitigate the negative impact of load shedding. We start our analysis with rural homes that cannot afford modern technologies against load shedding. Their only strategy is to reschedule home appliance usage outside of the load shedding period. For residential homes that can afford solar home systems such as solar panels and batteries, this study aims to prioritise both the sizing and operation of solar home technologies. For both cases, this study aims to optimise home appliance usage schedules amidst uncertain load shedding outages, to minimise inconvenience and loss of power supply. For this purpose, we consider a residential home with electrical appliances and power supply configurations as shown in Figure 3.1. At time k, the grid power supply is represented as $P_{gr}(k)$ and the residential load is represented as $P_L(k)$. The power output of the photovoltaic array is represented as $P_{PV}(k)$. The power flow to and from the battery bank is represented as $P_b(k)$. More precisely, the battery is charging when $P_b(k)$ is positive and discharging when $P_b(k)$ is negative. The load shedding outages are represented as $u_{LS}(k)$. The appliance switches to denote their usage schedules are represented as u_j , where $j = 1,2,...,8$. The index k refers to the

© University of Pretoria

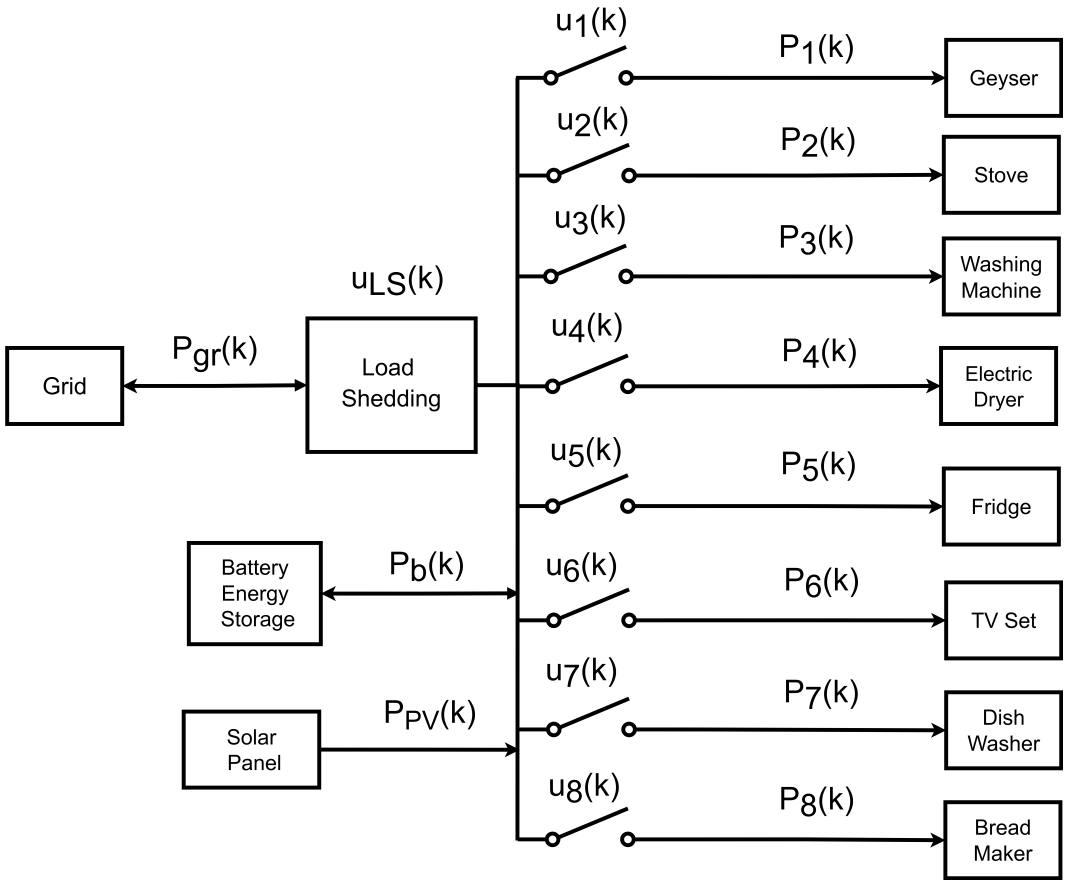

Figure 3.1. Illustrative schematic for home appliances control.

time period $[k\Delta t, (k+1)\Delta t)$, where Δt is the duration of the scheduling interval, with Δt = 10 minutes in this study.

In [29], home appliances are divided into fixed, flexible, and shiftable loads. Unlike the definition in [29] that the operational period of fixed loads could not be changed, some essential appliances such as refrigerators may be turned off involuntarily during load shedding. In this study, the operation schedule of the jth appliance is represented by the starting time S_j and the ending time E_j, respectively, with an operation duration of $D_j = E_j - S_j$.

In this study, home appliances are classified into four categories, namely continuous appliances (CA), adaptable appliances (AA), coordinated appliances (COA) and interruptible appliances (IA). The refrigerator is categorised as a continuous appliance due to the potential spoilage of food if it is turned off during an outage. Given the predetermined parameters S_j, E_j, and D_j within the interval $[S_j, E_j]$ for each appliance, adaptable appliances (AA) can be arbitrarily interrupted during their operation,

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

and unlike [42], the cumulative working time does not need to be satisfied to ensure the completion of tasks. The geyser is an adaptable appliance because it can be turned off temporarily without adversely affecting its operation. The stove is an adaptable appliance because its operation can be modified to allow a longer or shorter cooking duration. Coordination of operation between household appliances (COA) is essential because some appliances must operate sequentially to complete their tasks. The electric dryer should start its operation only after the washing machine has completed its laundry washing task. An interruptible appliance (IA), such as a dishwasher and a bread maker, can shift its working time to minimise the loss of power supply and inconvenience. However, when its operation starts, it must complete all its tasks without interruptions to its runtime.

3.1 LOAD SHEDDING MODEL

Load shedding is implemented nationally to protect the grid from total blackout. Blackouts occur when the demand for electricity is greater than the available supply, creating an unbalanced power system that causes grid collapse. Maintaining a balanced generation and demand is critical to prevent a national blackout, and when the national electricity grid is under pressure, Eskom is required to immediately shed the load, as agreed with the National Energy Regulator of South Africa (NERSA). The demand and generation with load shedding (Manual Load Reduction) is shown as a scatter plot in Figure 3.2.

The x axis of Figure 3.2 is the residual demand of the national grid from 2015 to 2023 of the Eskom dataset. The residual demand is the remaining demand after the national demand has been reduced by renewable generation. The y-axis of Figure 3.2 is the sum of dispatchable generation and manual load reduction (MLR). MLR is the demand that has been reduced due to load shedding. There is a strong, positive linear correlation between supply and load shedding and demand in Figure 3.2. Eskom has developed nine load shedding stages based on the required load that must be removed involuntarily from consumer load demand (MLR in the Eskom dataset), with each stage number representing the amount of gigawatts that must be removed from the grid, from 1 GW at stage 1 to 8 GW at stage 8. The relationship between MLR and load shedding stages is presented in Figure 3.3.

As future load shedding patterns can influence the solar home design and appliance schedules, we start to develop a model that is capable of characterising the dynamics of the load shedding patterns. The load shedding model in this study is derived from historical data on South Africa's load shedding

[2] www.eskom.co.za/dataportal/

© University of Pretoria

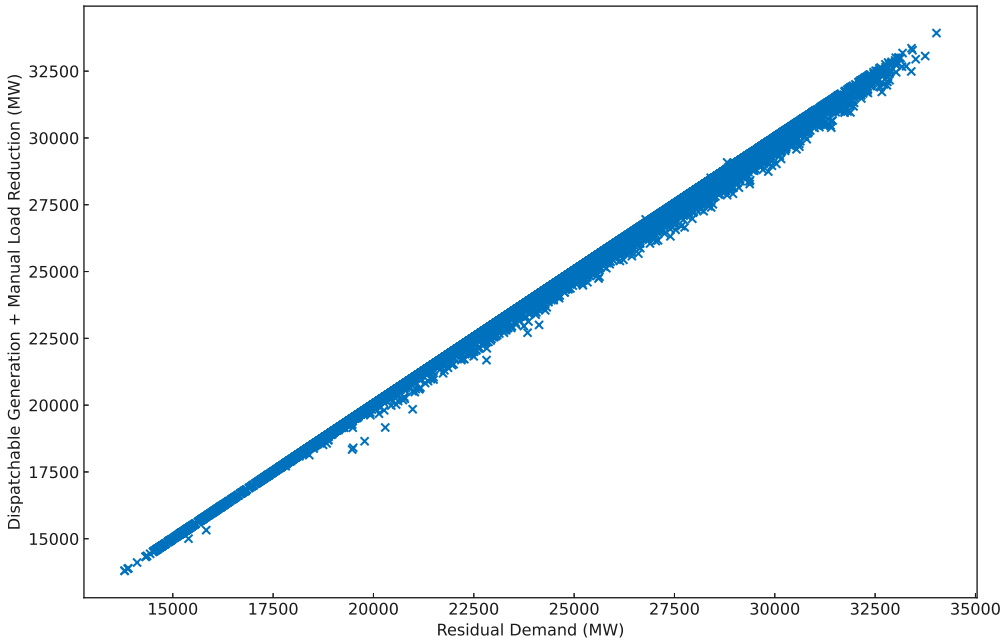

Figure 3.2. A scatter plot of dispatchable generation and load shedding, versus residual demand.

outages from 2015 to 2023. We first investigate whether there is a linear correlation between the features in the Eskom dataset, using a Pearson's correlation heat map in Figure 3.4. Each square in the heat map corresponds to the correlation coefficient between a pair of features, with the colour intensity inside each square indicating the magnitude and direction of the linear correlation. The colour gradient, which varies from purple to yellow, signifies the spectrum of correlation values. The purple colour represents a strong negative correlation (close to -1), suggesting that as one feature increases, the other tends to decrease. In contrast, yellow symbolises a strong positive correlation (close to 1), indicating that both features are increasing or decreasing together. Shades closer to blue represent little to no correlation (values near 0), denoting that changes in one feature do not predict changes in the other. The yellow block in Figure 3.4 that contains residual forecast, RSA contracted forecast, dispatchable generation, residual demand, and RSA contracted demand indicates a strong positive linear correlation between these features. This means that there is a strong positive linear correlation between dispatchable generation and both forecast and true demand, which is an important observation because during load shedding the demand is larger than the supply and the objective from a utility perspective is to reduce the demand to match the available generation. However, this strong linear correlation is only evident inside the deep yellow block. It is much more challenging to determine the

© University of Pretoria

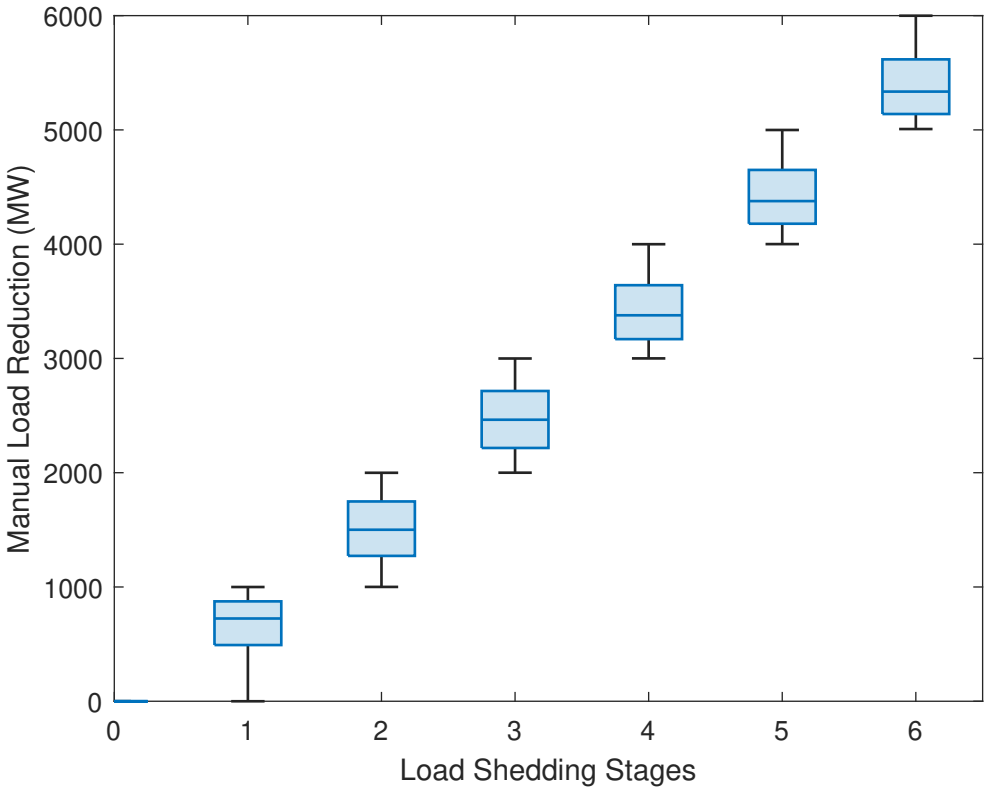

Figure 3.3. A box and whisker plot on load shedding stages and manual load reduction (load shedding outages) on Eskom's dataset.

stages of load shedding (in the fifth or last row) when both the national stages and the regional stages do not have strong Pearson's correlation coefficients with the other features.

The Minimum Redundancy Maximum Relevance (MRMR) algorithm is a feature selection method used to identify a set of features that are highly relevant to a target and minimally redundant with respect to each other. In doing so, MRMR addresses some of the limitations associated with Pearson's correlation coefficient. Figure 3.5 shows the importance scores of the features sorted using the MRMR algorithm. Unlike Pearson's correlation, which specifically measures linear relationships, MRMR can capture both linear and nonlinear relationships between features and the target of predicting load shedding stages. This is especially useful when the relationship between features and the target is complex and not strictly linear, as is evident with load shedding stages. One of the key principles of MRMR is to select features that complement each other's information about the target. This means that it considers multivariate interactions, ensuring that the selected features together provide a

© University of Pretoria

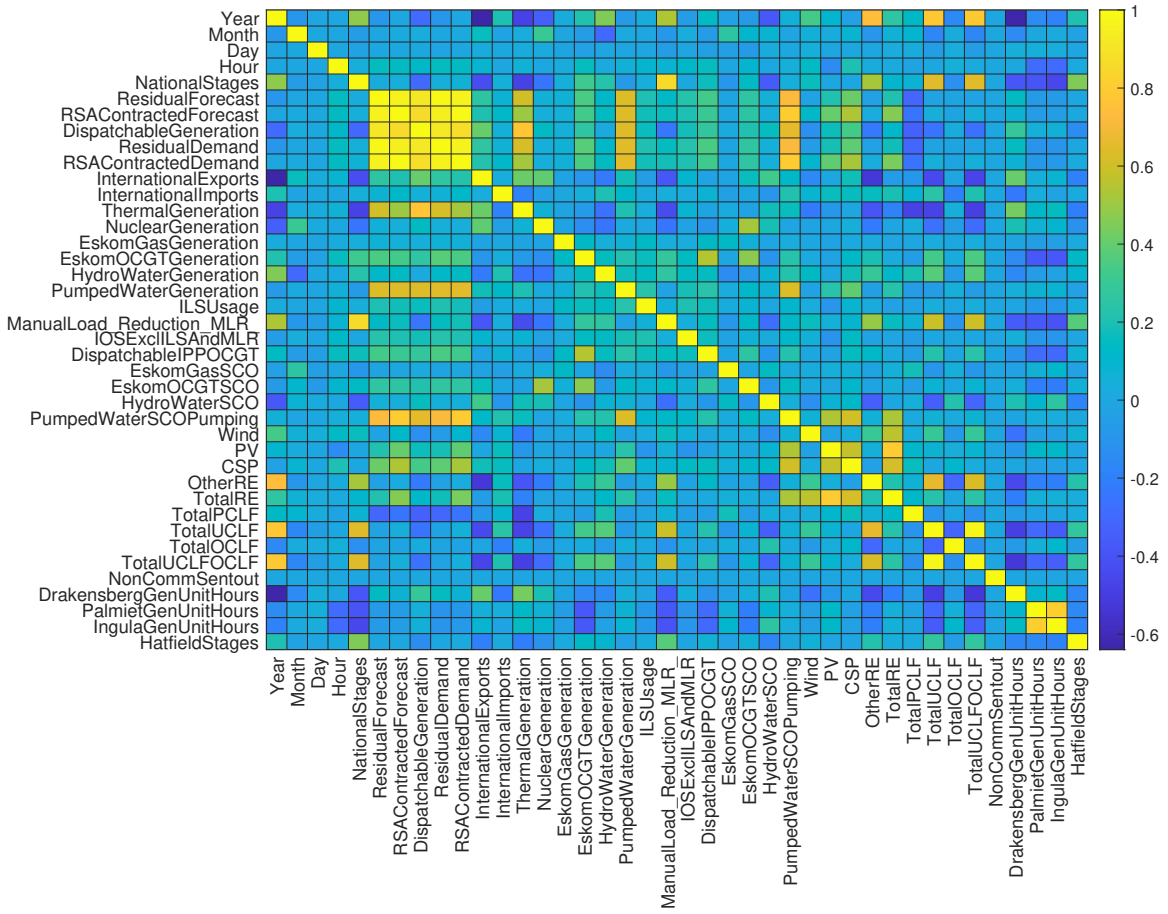

Figure 3.4. A heat map of Pearson's correlation coefficients on Eskom's dataset.

comprehensive understanding of the target. The maximum relevance of the MRMR algorithm ensures that the selected features are highly relevant to the target, addressing the challenge of selecting features that might have a high correlation with each other but low relevance to the target. From the feature ranking in Figure 3.5, it is evident that the three main features to predict load shedding stages are month, hour, and year. As a result, the four features that are selected for the load shedding model are month, hour, year, and day. Although the day feature is ranked as only the 16^{th} most relevant feature, the temporal features of month, hour, and year suggest that time-based attributes are highly relevant to predict load shedding.

The K-Nearest Neighbours (KNN) is a nonparametric machine learning algorithm that is used for both classification and prediction tasks. In this study, it is used to predict the load shedding stages. The core idea behind KNN is to classify or predict the value of a new data point based on its proximity to its nearest neighbours K in the training dataset. The Mahalanobis KNN used in this study is a variant

© University of Pretoria

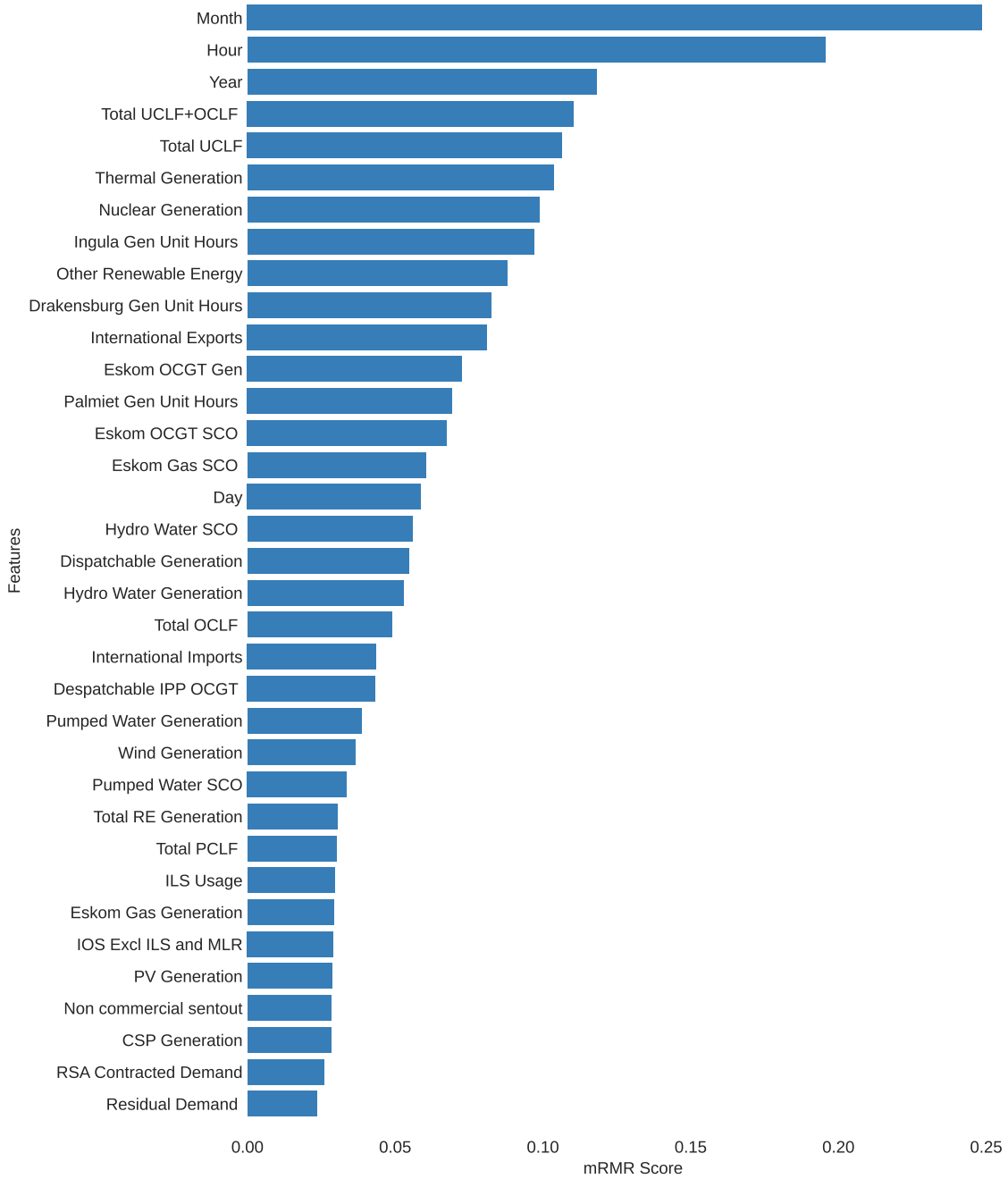

Figure 3.5. Feature importance scores when predicting load shedding stages sorted using a Minimum Reduncancy Maximum Relevance algorithm in MATLAB on Eskom's dataset.

of the KNN algorithm that takes into account the covariance structure of the data. It considers both the distances between feature vectors and their covariance when finding the nearest neighbours. The Mahalanobis KNN offers alternatives to the traditional KNN algorithm by incorporating additional information about the covariance structure or rank-order relationships of the features. The output are

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

the load shedding stages at specified future hours.

The KNN algorithm is based on the idea that the nearest neighbours to an unknown load shedding pattern at time k, provide important classification information. The Mahalanobis KNN is used to maximise the classification of a time instance k initially formulated in [43], to the most common class of its nearest neighbours,

$$x_{LS}(k) = \underset{x_{LS} \in C}{\operatorname{argmax}} \sum_{n=1}^{K} \delta(x_{LS}, x_{LS}(y_n)), \tag{3.1}$$

where $\mathrm{x}_{LS}(k)$ is the predicted load shedding stage at time k and is an integer variable that can only take on values of the class C, which is the set of integer class labels representing the nine different stages of load shedding, even though only seven have been implemented between 2015 and 2023. Each $\mathrm{x}_{LS}(k)$ value correlates to the number of gigawatts of load demand that must be removed from the grid, ranging with 1 GW of load demand that is involuntarily shed from end users when $\mathrm{x}_{LS}(k) = 1$, to 6 GW that is shed when $\mathrm{x}_{LS}(k) = 6$. K is the number of nearest neighbours at time k, y_n are the nearest neighbours at time k, with both variables used to determine how many of the closest data points are considered when making the prediction, and $\delta(x_{LS}, x_{LS}(y_n)) = 1$ if $x_{LS} = x_{LS}(y_n)$ and $\delta(x_{LS}, x_{LS}(y_n)) = 0$ otherwise. $\delta(x_{LS}, x_{LS}(y_n))$ is a binary function that is essential for counting the occurrences of each stage among the nearest neighbours. In this study, (3.1) determines the most common load shedding stage among the nearest neighbours K of a given time instance k in the dataset from 2015 to 2022, and the class that occurs the most frequently among these neighbours is predicted as the load shedding stage $\mathrm{x}_{LS}(k)$. This approach is typical of KNN algorithms, where the prediction is based on the majority vote of the nearest neighbours.

The performance of the KNN in classification is hindered by using a fixed K value for all queries in the search phase [44] and using a fixed distance metric to quantify the distances between a new data point and the training dataset. These are often arbitrarily chosen, and this has a negative impact on the performance of the KNN. In this study, Bayesian hyperparameter optimisation with 300 iterations and an Expected Improvement acquisition function is used to systematically determine the number of neighbours and the distance metric. To evaluate the performance of the classification model, the accuracy formulated in [45] is defined as

$$Accuracy = \frac{CP + CN}{CP + CN + WP + WN}, \tag{3.2}$$

where CP are the correct positive classifications, CN are the correct negative classifications, WP are the wrong positive classifications and WN are the wrong negative classifications. Furthermore, the positive

© University of Pretoria

predictive values (PPV), which is the proportion of correctly classified observations per predicted class, and the False Discovery Rates (FDR), which is the proportion of incorrectly classified observations per predicted class, are presented in the confusion matrix in Figure 3.6, respectively.

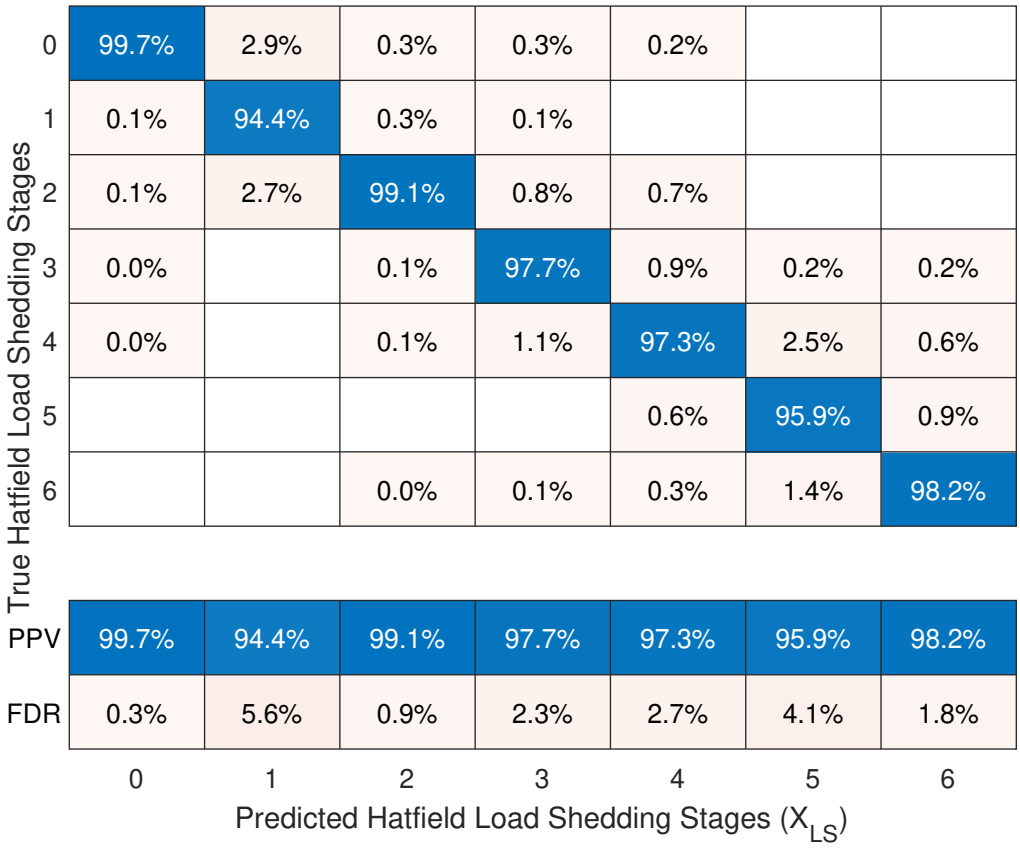

Figure 3.6. The confusion chart of Mahalanobis KNN predictions of the load shedding stages in Hatfield, South Africa.

A unity binary power supply means that there is no power outage and that the grid can meet the load demand.

$$u_{LS}(k) = \begin{cases} 1, & \text{if } x_{LS}(k) = 0, \text{ no power outage} \\ 0, & \text{if } x_{LS}(k) > 0, \text{ power outage} \end{cases} \tag{3.3}$$

where $u_{LS}(k)$ formulates the general power outage schedule at time k. This novel time series classification model enables a determination of load shedding outage times and load shedding outage duration in South Africa.

The novel load shedding model in (3.3) can be used with any area's power outage schedule. To translate the predicted stages of load shedding into power outages and the likelihood of a power outage at any

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

given time, k, the loss of power supply (LPS) at time k, is defined as

$$LPS(k) = P_L(k)\Delta t - P_{gr}(k)u_{LS}(k)\Delta t, \tag{3.4}$$

where $P_L(k)$ is the schedulable load, and $P_{gr}(k)$ is the decision variable representing the power from the grid. Loss of power supply probability (LPSP) measures the probability of a power outage in which the load is not satisfied due to load shedding and is formulated as

$$LPSP(k) = \frac{\sum_{k=1}^{N} LPS(k)}{\sum_{k=1}^{N} P_L(k)\Delta t}, \tag{3.5}$$

where $N = 144$ is the total number of sampling periods. The training and testing dataset are split into a 90/10 ratio from the total entries. The testing dataset is not used in the KNN training, but is reserved for evaluating the performance of the KNN, after the training and the 5-fold cross-validation. The dataset has the following characteristics:

- Total entries: 74 474;
- Training entries: 65 714;
- Training dates: 1 January 2015 - 31 December 2022;
- Testing entries: 8 760;
- Testing dates: 1 January 2023 - 31 December 2023.

The accuracy of the KNN predicted Hatfield stages of load shedding, x_{LS}, is compared to the accuracy of other prediction methods, namely a support vector machine (SVM), an artificial neural network (ANN), and a Decision Tree (Tree), and the comparison results are provided in Table 3.1. The same Bayesian hyperparameter optimisation methodology used for the KNN is used for the ANN, SVM and Tree, not only to improve the accuracy of each prediction method but also to create a fair comparison among all the methods. The KNN has the highest accuracy in forecasting each of the load shedding stages and is highlighted in blue.

The inherent inability to explain how the KNN learns from the 65 714 training entries to make a prediction on each of the 8 760 testing entries, despite its high-accuracy performance metric, makes it difficult to understand how the machine learning KNN determines the load shedding stage. A recent approach focussing on making machine learning models more explainable is the use of eXplainable Artifical Intelligence (XAI) to identify feature scores that quantify the significance of each feature for the prediction made by the model [46]. In this study, one of the two XAI methods comparatively analysed in [46] for feature selection, SHapley Additive exPlanations (SHAP), is used for model

© University of Pretoria

Table 3.1. The accuracy of different prediction methods when compared to the KNN, to forecast the load shedding stages in 2023, Hatfield, South Africa.

Stage	KNN (%)	SVM (%)	ANN (%)	Tree (%)
0	**99.7**	**99.7**	99.2	99.2
1	**94.4**	87.5	89.7	85.8
2	**99.1**	98.2	97.8	97.4
3	**97.7**	96.8	96.3	92.9
4	**97.3**	96.1	95.1	92.6
5	**95.9**	92.8	92.1	88.6
6	**98.2**	97.7	97.2	94.1

interpretability. The Shapley value of a feature for a prediction explains the deviation of that prediction from the average prediction. For a single-entry prediction, the sum of the Shapley values for all features corresponds to the total deviation of the prediction from the average.

In Figure 3.7, Shapley values are generated with KernelSHAP [47] to interpret the KNN model in MATLAB for a single-entry load shedding prediction. KernelSHAP is a model-agnostic weighted linear regression method used to interpret the predictions of any machine learning model. One of the key advantages of KernelSHAP is that it can be used with any machine learning model, including KNN, because it does not need to know how the model works internally, it only requires access to the model's predictions. The features of year 2023, month 8, day 29 and hour 15 are used with the KNN output predicting stage 4, which is the true load shedding stage. This study introduces a "confidence score" to interpret the Shapley values. The confidence score is not a measure of correctness; it is the level of confidence that the KNN has in its prediction of the load shedding stage, with a value of 1 explaining that the KNN is 100% confident in its prediction. The average confidence score for each load shedding stage in the 65 714 training entries is the probability that a randomly selected input from the testing entries would be predicted to belong to a specific stage by the KNN. Average confidence scores create a baseline to explain how confident the KNN is in making its prediction compared to whether it makes the correct prediction. Intuitively, we would expect the KNN to make an accurate prediction when it has a high confidence score and an inaccurate prediction when it has a low confidence score. The confidence scores enable us to trust the KNN output because they essentially inform us when the

© University of Pretoria

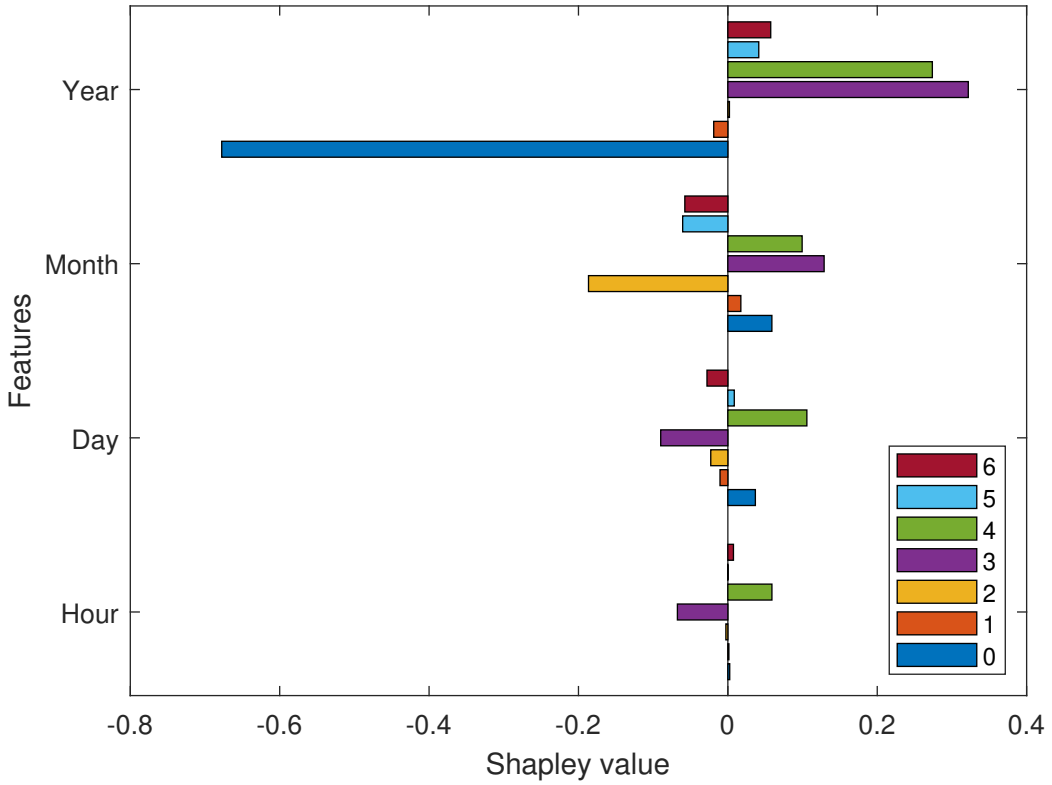

Figure 3.7. Shapley values to interpret a single load shedding prediction by the model for the KNN features of year 2023, month 8, day 29 and hour 15, with the KNN output predicting stage 4 which is the true load shedding stage.

machine learning model is not confident in its prediction, and we can intervene if necessary.

The average confidence scores for stages 0 to 6 for the four features are determined to be 0.6828, 0.0179, 0.2123, 0.0088, 0.0605, 0.0041, and 0.0135, respectively. The sum of the average confidence score and the Shapley values for the four features in predicting stage 0 (dark blue bar) is 0.0001. The KNN is therefore 0.01% confident in predicting that the features of year 2023, month 8, day 29 and hour 15 will result in no load shedding. Similarly, the KNN is 0.01% confident in predicting that the given features result in the load shedding stages 1, 2, 5 or 6. The low confidence score extends to the prediction of stage 3 which, although significantly higher than 0.01% (33.34%), is less than 50%. In explaining the confidence that the KNN has in predicting stage 4 of load shedding for the specific entry in Figure 3.7, the year feature adds a 0.3234 score to the average confidence score, the month feature adds a 0.0952 score to the average confidence score, the day feature adds a 0.0939 score to the average

© University of Pretoria

confidence score, and the hour feature adds a 0.0936 score to the average confidence score. The KNN is, therefore, 66.66% confident with its prediction that the features of year 2023, month 8, day 29 and hour 15 will result in stage 4 load shedding, which is the true stage of load shedding for that specific prediction.

After the KNN accurately and confidently forecasts the load shedding stages for each of the 8 760 hours in 2023, the forecast is analysed statistically. There are multiple load shedding outages during a single day of load shedding, which can last longer than 2 hours per outage. The average cumulative duration of load shedding outages over a 24-hour period in 2023 is shown in Figure 3.8 as 7.5 hours.

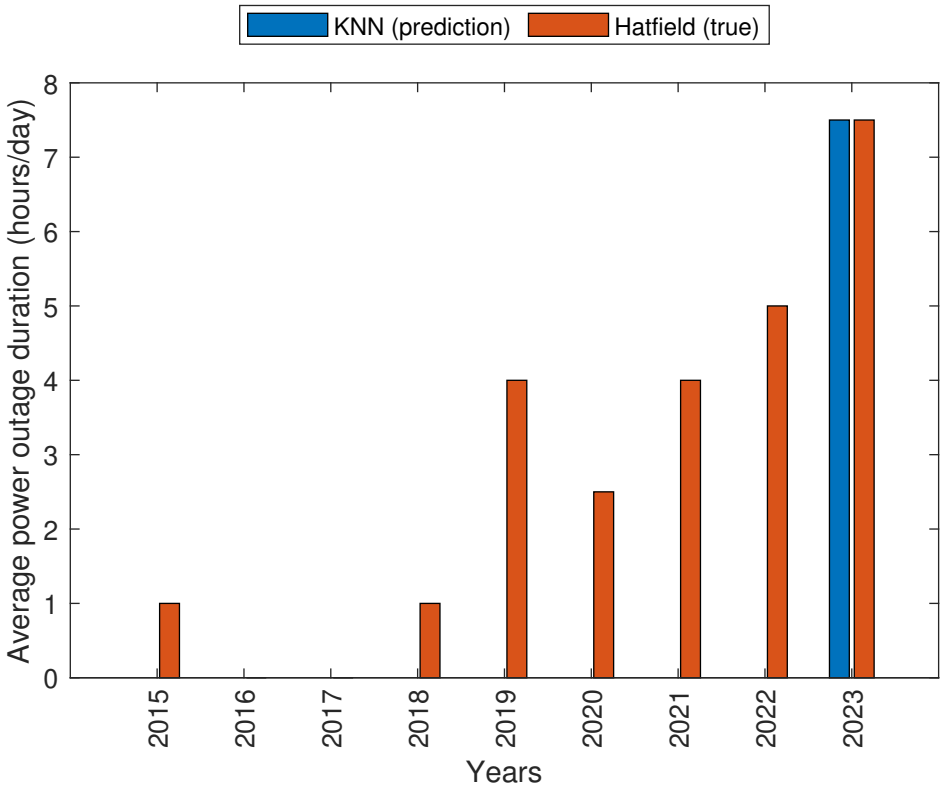

Figure 3.8. The average load shedding hours per day in Hatfield, South Africa.

3.2 PV SYSTEM MODEL

The PV system consists of a combination of several solar cells that convert solar energy into direct current power. The maximum hourly output power (kW) of the PV system, $P_{PVA}(k)$ is formulated in [48] as

$$P_{PVA}(k) = \eta_{PV} I_{PV}(k) A_c, \tag{3.6}$$

© University of Pretoria

where η_{PV} is the efficiency of the solar PV generator, $I_{PV}(k)$ is the solar irradiation (kWh/m^2) incident on the PV panels in the kth time interval, and A_c is the surface area of the given solar panels (m^2). In this study, A_c is a sizing decision variable determined in the optimisation algorithm. The lower and upper bounds of the cross-sectional surface area sizing decision variable are subject to the available roof space of the house.

3.3 BATTERY ENERGY STORAGE SYSTEM MODEL

The battery energy storage system in Figure 3.1 can be formulated as

$$SOE_b(k) = SOE_b(0) + \eta \sum_{\tau=0}^{k-1} \text{sgn}\left[P_b(\tau)\right] \Delta t, \tag{3.7}$$

$$\eta = \begin{cases} \eta_C, & \text{if } P_b(\tau) > 0 \\ \eta_D, & \text{if } P_b(\tau) < 0 \end{cases} \tag{3.8}$$

where P_b is the battery power, n_C is the converter charging efficiency, n_D is the converter discharging efficiency and SOE_b is the battery's state-of-energy (SOE) measured in kWh. The initial state of energy is denoted as $SOE_b(0)$ and is assumed to be 60% of the battery capacity C_b. The operational constraints of the battery energy storage system should include the lower and upper boundaries on the battery,

$$SOE_b^{min} \leq SOE_b(k) \leq C_b, \tag{3.9}$$

which ensure that the battery does not discharge below its minimum capacity SOE_b^{min} and does not charge beyond its maximum rating. The C_b is a sizing decision variable and can be determined in the optimisation algorithm. The minimum SOE of the battery is constrained by the depth of discharge (DOD) to mitigate against the degradation of the battery energy storage system if it were fully discharged,

$$SOE_b^{min} = (1 - DOD)C_b. \tag{3.10}$$

3.4 LOSS OF POWER SUPPLY

The operation and the capacity optimisation of the solar home system is formulated as an optimisation problem in this section. The first objective is to minimise the loss of power supply represented as

$$J_{LPS} = \sum_{j=1}^{8} \sum_{k=1}^{N} \left[P_j u_j^{\text{opt}}(k) - P_{gr}(k) u_{LS}(k) \right] \Delta t, \tag{3.11}$$

where $u_j^{opt}(k)$ represents the optimised binary switching status of appliance j at time k, P_j is the rated power of the jth appliance and the objective J_{LPS} is the loss of power supply due to load shedding.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

3.5 INCONVENIENCE BY LOAD SHEDDING

Homes are inconvenienced during load shedding when they have to change their daily habit of appliance usage. In this study, we define the daily habit of appliance usage into a baseline load profile using a bottom-up approach. During load shedding, people have to change their energy usage behaviour. The duration when the baseline operation is different from the actual rescheduled appliance usage profile is defined as "inconvenience" in this study. For example, if the inconvenience is 2 hours, then it implies that one of the appliances have been switched on or off for 2 hours with a different schedule from the baseline operation of the same appliance. We assume that this schedule change is involuntary and caused by load shedding. There is an inherent, uncoordinated, scheduling of appliances by consumers in response to load shedding, that may differ from the baseline studied in [29], where DSM participation was voluntary. The assumption made in this study is that the baseline during involuntary load shedding is $P_{baseline}(k) \times u_{LS}(k)$.

The second objective is to minimise the load shedding inconvenience formulated as

$$J_I = \sum_{j=1}^{8} \sum_{k=1}^{N} \left[u_j^{bl}(k) u_{LS}(k) - u_j^{opt}(k) \right]^2 \Delta t, \tag{3.12}$$

where $u_j^{bl}(k)$ represents the baseline operating status of the jth appliance.

3.6 HOME APPLIANCES SCHEDULING MODEL

A residential household includes various appliances, and the purpose of scheduling is to find the optimal operations for each appliance while considering a user's inconvenience index, during the various stages of load shedding as formulated as

$$J_I = J_I^{\mathrm{AA}} + J_I^{\mathrm{COA}} + J_I^{\mathrm{CA}} + J_I^{\mathrm{IA}}. \tag{3.13}$$

3.7 DECISION VARIABLES

This study considers both the sizing optimisation problem and the operation optimisation problem. The decision variables include the power flows within the system, the scheduling of home appliances, and the sizing of the alternative energy sources. The decision variables are arranged in a single (2 + 11N)×1 vector,

$$\mathbf{X} = \begin{bmatrix} \mathbf{V}^T & \mathbf{U}^T & \mathbf{P}^T \end{bmatrix}^T_{(2+11N)\times 1}, \tag{3.14}$$

where

$$\mathbf{V} = \begin{bmatrix} A_c & C_b \end{bmatrix}^T, \tag{3.15}$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

$$\mathbf{U} = \left[u_1^{opt}(1) \ \ldots \ u_1^{opt}(N) \ \ldots \ u_8^{opt}(1) \ \ldots \ u_8^{opt}(N)\right]^T_{8N\times 1}, \tag{3.16}$$

$$\mathbf{P} = \left[P_{gr}(1) \ \ldots \ P_{gr}(N) \ P_{PV}(1) \ \ldots \ P_{PV}(N) \ P_b(1) \ \ldots \ P_b(N)\right]^T_{3N\times 1}. \tag{3.17}$$

3.8 CONSTRAINTS

The novel optimisation problem is subject to constraints on the operation of the system given in (3.6), (3.9) and the system capacity given as

$$A_c^{min} \leq A_c \leq A_c^{max}, \tag{3.18}$$

with the lower and upper bounds referring to the minimum and maximum available residential roof space for the solar panel installation. These specifications can be found in Table 4.1. The CA constraints described in [19] are improved to consider load shedding,

$$\sum_{k=S_5}^{E_5} u_5^{opt}(k) = D_5, \tag{3.19}$$

$$\sum_{k=S_6}^{E_6} u_6^{\text{opt}}(k) = D_6, \tag{3.20}$$

where D_5 and D_6 are the operation duration of appliance 5 and 6, respectively. The AA loads are constrained by

$$\sum_{k=1}^{N} u_{LS}(k) u_2^{opt}(k) = D_2 + g_2, \tag{3.21}$$

where g_2 is the additional operation period chosen for appliance 2. This constraint offers flexibility to users, as they can adapt (increase or decrease) the operation period wherever possible and convenient. For example, if the estimated cooking duration is 30 minutes, the user can choose to increase or decrease it by g_2. The COA loads are constrained by

$$\sum_{k=1}^{N} u_{LS}(k) u_j^{opt}(k) = D_j, \quad j = 3,4. \tag{3.22}$$

The coordination of operation for COA loads is sequential. To account for this sequential operation [29], the formulation is

$$S_j + D_j \leq D_q, \quad q \neq j \tag{3.23}$$

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

which indicates the qth appliance can only start its operation after the jth appliance has completed its task. The time the jth appliance starts and stops its operation is represented by S_j and E_j, respectively, with an operation duration of $D_j = E_j - S_j$. The IA loads are constrained by

$$\sum_{k=S_j}^{E_j} u_{LS}(k)\, u_j^{opt}(k) \geq D_j, \quad j = 7,8. \tag{3.24}$$

The total power consumed by the home appliances described in [49] is given by

$$P_L(k) = \sum_{j=1}^{N} P_j u_j^{\text{opt}}(k), \tag{3.25}$$

and the power balance equality constraint is given as

$$P_L(k) = P_{PV}(k) + P_b(k) + P_{gr}(k) u_{LS}(k), \tag{3.26}$$

which ensures that the power demands of the loads and the battery relative to the inconvenience index, at any time k, should be met through the various stages of load shedding. The PV system model in (3.6) is constrained by

$$P_{PV}(k) + P_b(k) \leq P_{PVA}(k) + P_{gr}(k) u_{LS}(k). \tag{3.27}$$

3.9 OPTIMISATION ALGORITHM

The proposed model is a multi-objective mixed integer nonlinear programming (MOMINLP) problem, which is solved by the SCIP algorithm from the OPTI toolbox interfaced with MATLAB. SCIP is currently one of the fastest noncommercial solvers available for MOMINLP. To use the SCIP algorithm, the developed multi-objective optimisation problem is first converted into a single-objective problem using a weighted sum method [50], which leads to the following combined objective function

$$J = w_1 J_{LPS} + w_2 J_I, \tag{3.28}$$

where w_1 and w_2 are the loss of power supply and the inconvenience weighting factors, respectively, and $w_1 + w_2 = 1$. The settings of the SCIP solver used in this study are given in Table 3.2.

[3] www.controlengineering.co.nz

© University of Pretoria

Table 3.2. SCIP settings.

Solver	SCIP
Maximum iterations	1500
Maximum nodes	100 000
Maximum run-time	300 seconds
Tolerance of the objective function's value	$1e^{-7}$

© University of Pretoria

CHAPTER 4 CASE STUDY: EVALUATING STRATEGIES TO MINIMIZE THE IMPACT OF LOAD SHEDDING ON A WORKING-CLASS HOUSEHOLD

A typical working-class household in South Africa is chosen to evaluate the four strategies that minimise the loss of power supply and the inconvenience caused by load shedding. It is assumed that the household can install batteries or solar panels to provide self-generated power when load shedding occurs. Without these solutions, this home suffers from power outages and reschedules its usual use of appliances to periods outside load shedding. The power usage of the home is optimised over a 24-hour period with a sampling interval of 10 minutes and the total sampling interval N=144.

The load shedding outage profile based on the average load shedding outage duration and on the hourly load shedding probability predictions for 2023 is shown in Figure 4.1 and is used in the case study. When the binary indicator $u_{LS}(k) = 1$, the house has power supply; otherwise, it suffers from load shedding. Outages occur in the periods with the highest probability of load shedding, based on KNN predictions. The specifications of the hybrid system considered are shown in Table 4.1 and the occupant's habitual operation of the appliances is given in Table 4.2, which shows the duration of each appliance, D_j, the average power consumption of each appliance, P_j, as well as the start and end of the scheduled working time interval of each appliance based on the preferred power usage of the residents (baseline), represented by S_j and E_j, respectively. These eight appliances are representative of the commonly used household devices as reported in [29, 38]. The baseline energy consumption (Power demand w/o LS) in Figure 4.2 (a) portrays two peaks: the morning peak and the evening peak. If the appliances operate only on the grid, the energy consumption for 24 hours based on the baseline schedule is 28.08 kWh.

© University of Pretoria

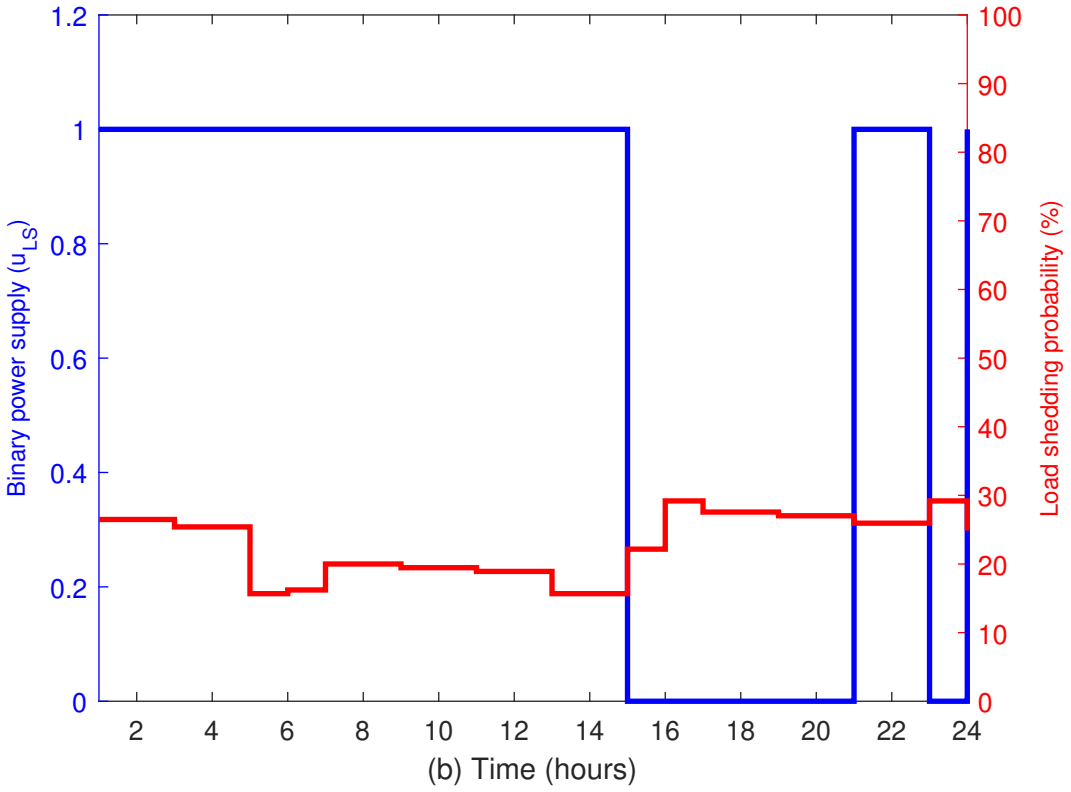

Figure 4.1. The load shedding outage profile based on the average load shedding outage duration and on the hourly load shedding probability predictions for 2023.

Table 4.1. Specifications of the hybrid power system.

Maximum roof space	20 m^2
Battery's depth of discharge	50%
Battery charge efficiency	85%
Battery discharge efficiency	100%

The results shown in Table 4.4 show that the household experiences a loss of power supply of 16.23 kWh during the average load shedding events in 2023. If no actions are taken, this household is inconvenienced for 1 250 minutes. The effects of not taking any action during load shedding over a week are shown in Table 4.3 with varying stages of load shedding. The household can experience more than 55 hours of inconvenience when no action is taken for a week of stage 8 load shedding.

For households that can implement measures against load shedding, we evaluated the results of

© University of Pretoria

Table 4.2. Appliance specifications and baseline energy usage scheduling without load shedding.

j	Appliances	P_j (kW)	Duration	Baseline		Classification
			D_j (min)	S_j	E_j	
1	Geyser	3.0	120	31	42	AA
			120	104	115	AA
2	Stove	2.5	30	32	34	AA
			50	113	117	AA
3	Washing machine	0.5	60	109	114	COA
4	Electric dryer	2.0	30	116	118	COA
5	Refrigerator	0.1	1440	1	144	CA
6	TV set	0.2	180	104	121	CA
7	Dishwasher	1.8	150	116	130	IA
8	Bread maker	1.5	150	118	132	IA

Table 4.3. Weekly effects of South Africa's load shedding stages when no action is taken.

Stage	Demand (kWh)	Grid supply (kWh)	Loss of power supply (kWh)	Inconvenience (hours)
0	196.56	196.56	0	0
1	196.56	185.00	11.56	5.95
2	196.56	171.07	25.49	14.55
3	196.56	159.26	37.299	22.44
4	196.56	143.39	53.17	32.12
5	196.56	132.46	64.10	37.85
6	196.56	121.52	75.47	43.59
7	196.56	111.58	84.98	48.96
8	196.56	99.36	97.20	55.41

four different strategies depicted in Figures 4.2-4.7. These strategies, which vary according to their investment options, include: appliance scheduling only, appliance scheduling with solar panels, appliance scheduling with batteries, and appliance scheduling with solar panels and batteries. The following evaluation aims to determine the most suitable scheduling and sizing strategy to combat load

© University of Pretoria

shedding. All strategies reported in the case study use the KNN prediction of the average duration of load shedding to reschedule appliances outside of the high-risk load shedding periods.

Table 4.4. The daily loss of power supply, inconvenience, sizing and energy management for all four strategies.

Strategies	J_{LPS} (kWh)	J_I (min)	A_c (m^2)	C_b (kWh)	P_{gr} (kWh)	P_{PV} (kWh)	$\lvert P_b \rvert$ (kWh)
No action + LS	16.23	1250	0	0	11.85	0	0
HAS+LS	1.35	1810	0	0	26.73	0	0
HAS+LS+PV	0.12	1810	2.58	0	26.37	2.83	0
HAS+LS+BESS	0	100	0	9.6	18.3	0	9.78
HAS+LS+PV+BESS	0	100	5.16	9.6	12.28	5.65	15.81

4.1 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING

The first strategy considers only HAS as a response to load shedding. It assumes that appliance scheduling is the only tool available to the consumer to mitigate against the effects of load shedding. The loss of power supply and the inconvenience caused by load shedding, subject to the constraints in Section 3.8, are minimised in a MOMINLP. It is shown that scheduling appliances without any backup power reduces the loss of power supply by 91.7%. However, the inconvenience increases by 44.8% as observed in Table 4.4.

The periods and duration of inconvenience for all appliances and the optimal schedule for rescheduling using this strategy are shown in Figure 4.2. The duration when the baseline operation in Figure 4.2 (a) (Power demand w/o LS) is different from the actual rescheduled appliance usage profile in Fig. 4.2 (a) (Optimal demand with HAS + LS) is shown as the inconvenience in Figure 4.2 (c). The increased grid power supply is shown in Figure 4.2 (b). To significantly reduce the loss of power supply without any alternative energy source, households are severely inconvenienced. This optimal strategy has the least desirable minimisation of both inconvenience and loss of power supply of the four optimal strategies considered. The HAS+LS strategy offers the most suitable solution for households that do not have the ability to purchase solar panels or batteries, because it reduces the loss of power supply by 91.7% when compared to not taking action during load shedding.

© University of Pretoria

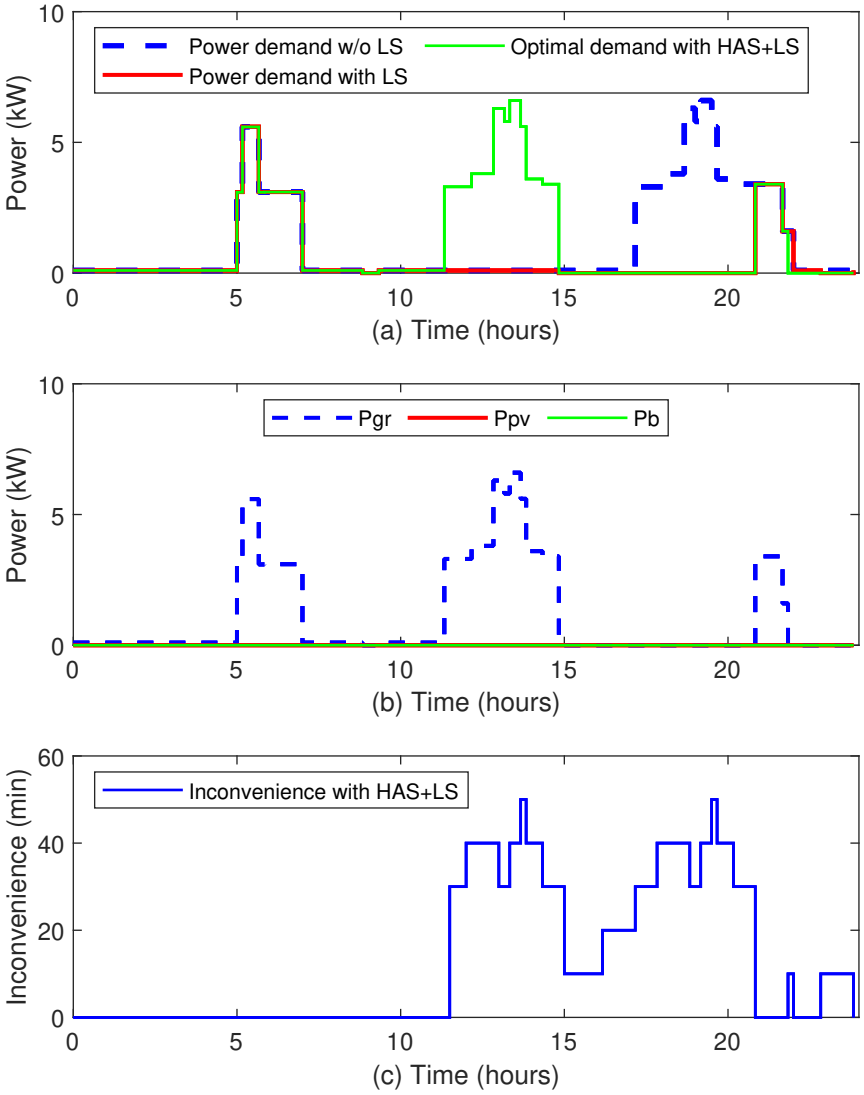

Figure 4.2. (a) The optimal power demand for scheduling with load shedding; (b) with the optimal power flow management of the energy sources; (c) and the inconvenience.

4.2 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING AND A PHOTOVOLTAIC SOLAR PANEL ARRAY

The second strategy considers HAS and the introduction of an alternative energy source, in the form of a photovoltaic solar panel array. The A_c is the sizing decision variable in (3.6) determined in the optimisation, subject to the bounds in (3.18) and the inequality constraint in (3.27). The LPS for the second case study is modified from (3.12) to (4.1)

© University of Pretoria

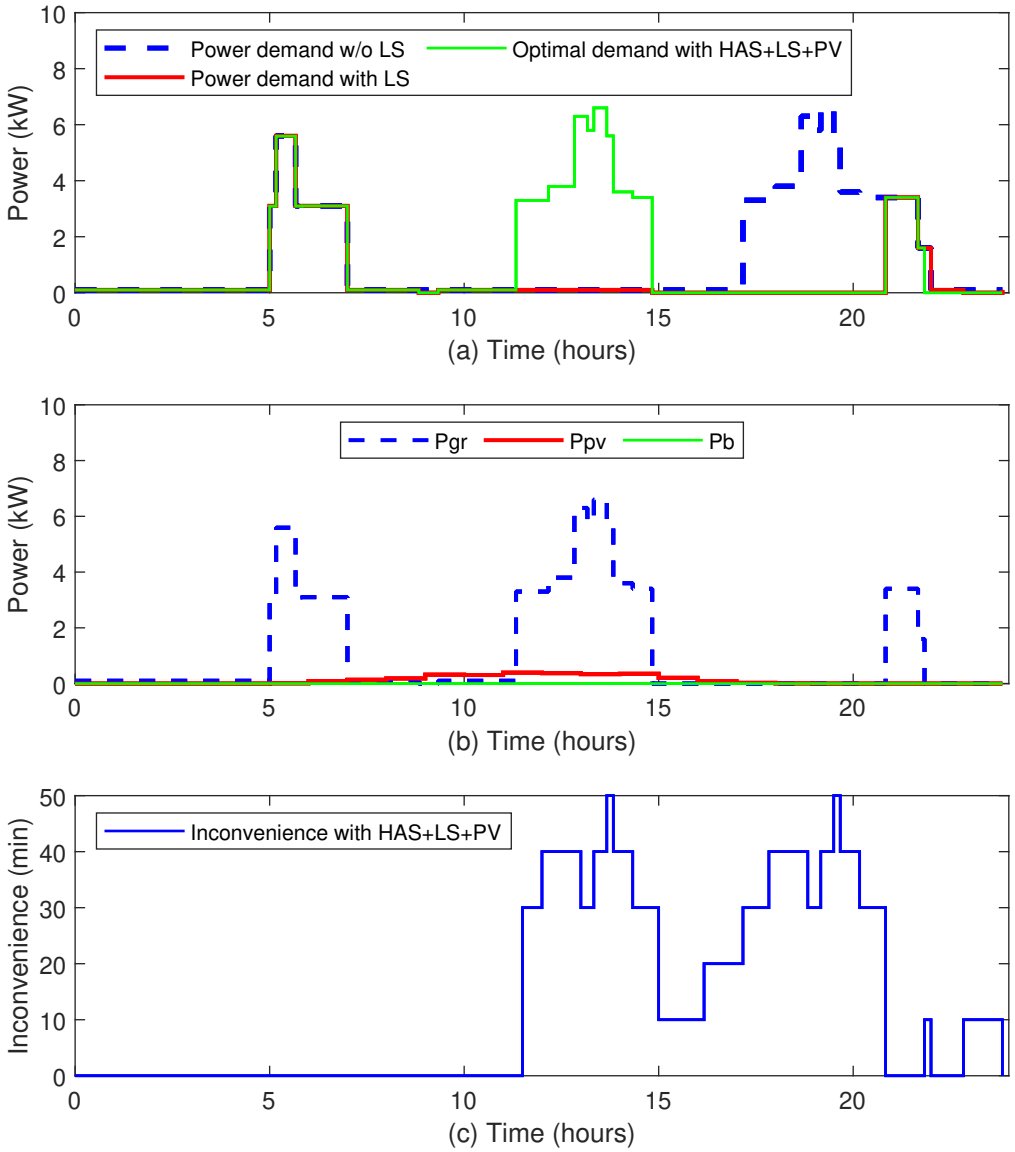

Figure 4.3. (a) The optimal power demand for scheduling with load shedding using Figure 4.1, and a photovoltaic solar panel array (b) with the optimal power flow management of the energy sources (c) and the inconvenience.

$$J_{LPS} = \sum_{j=1}^{8} \sum_{k=1}^{N} \left[P_j u_j^{\text{opt}}(k) - P_{gr}(k) u_{LS}(k) - P_{PV}(k) \right] \Delta t. \tag{4.1}$$

© University of Pretoria

The inconvenience caused by load shedding, subject to the constraints in Section 3.8, is minimised in a MOMINLP with $w_1 = 0.5$ and $w_2 = 0.5$. Appliance scheduling with an optimally sized solar panel array of 0.391 kWp, reduces the loss of power supply by 96% and increases the inconvenience. The reason the sizing of the panel array as formulated in (3.6) and constrained by (3.18) results in a relatively small solar panel capacity is because the average probability of load shedding during the evening is significantly higher than during the day. The power generated by the solar panel array in Figure 4.3 (b) is only generated during the day. The optimal schedule for rescheduling is shown in Figure 4.3 (a) and the periods and duration of inconvenience using this strategy are shown in Figure 4.3 (c). Considering the limitations of finding a 0.391 kWp panel, as determined through optimisation, a single 390 Wp standard panel is recommended for this strategy. Although the HAS+LS+PV strategy suffers the same severe inconvenience as the HAS+LS strategy, it minimises the loss of power supply by 4.7% more. Government incentives for solar panels provide a strong motivation to consider this strategy.

4.3 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING AND BATTERY ENERGY STORAGE

The third strategy considers HAS and the introduction of a battery energy storage. The C_b is the sizing decision variable in (3.7) determined in the optimisation, subject to the inequality constraints in (3.9) and the bounds in (3.10). The LPS for the third case study is described as

$$J_{LPS} = \sum_{j=1}^{8} \sum_{k=1}^{N} \left[P_j u_j^{\text{opt}}(k) - P_{gr}(k) u_{LS}(k) - P_b(k) \right] \Delta t. \tag{4.2}$$

The inconvenience and loss of power supply caused by load shedding, subject to the constraints in Section 3.8, are minimised with the weighting factor $w_1 = 0.5$ and $w_2 = 0.5$. The resulting appliance scheduling with an optimally sized battery capacity of 9.6 kWh eliminates the loss of power supply and reduces the inconvenience by 92%. This is one of two strategies that eliminates the loss of power supply, and the other strategy (HAS+LS+PV+BESS) also involves a battery capacity of 9.6 kWh, as observed in Table 4.4. The Pareto set of solutions for sizing batteries for scheduling appliances with the grid is shown in Figure 4.5 (a), and provides insight into the trade-off between inconvenience and loss of power supply with battery capacity. The idea of Pareto optimality is often used to describe solutions for multi-objective optimisation problems. A solution point is Pareto optimal if it is not possible to move from that point in the feasible region and improve at least one of the objectives in (3.28) without affecting the other. A point is weakly Pareto optimal if it is not possible to move from that point in the feasible region and minimise both inconvenience and loss of power supply simultaneously. In analysing Figure 4.5 (a) more closely, the Pareto set of solutions for the size of

© University of Pretoria

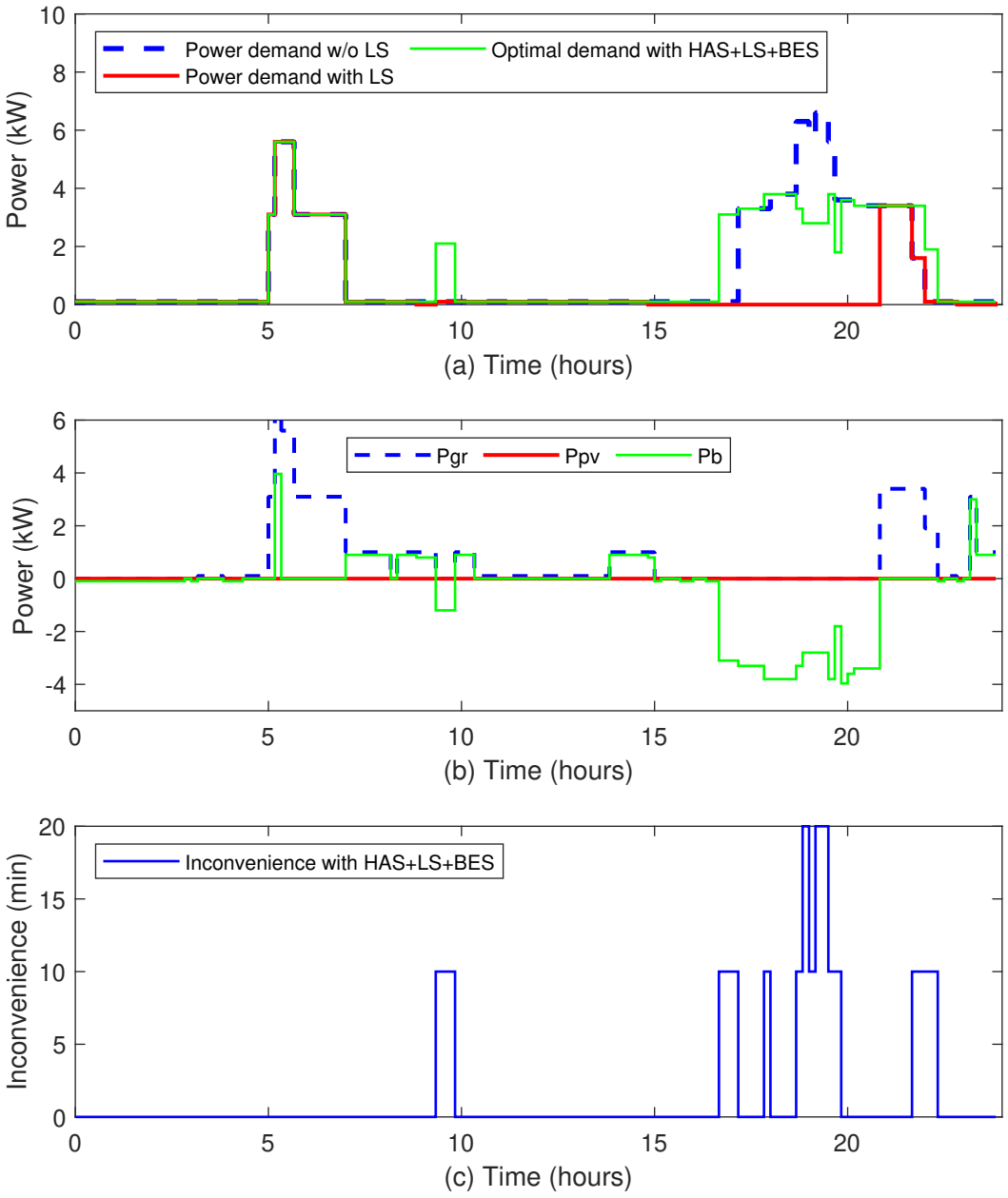

Figure 4.4. (a) The optimal power demand for scheduling with load shedding using Figure 4.1, and battery energy storage, (b) with the optimal power flow management of the energy sources (c) and the inconvenience.

batteries, to minimise the inconvenience caused by load shedding, is always on the boundary of the feasible region in Figure 4.6, specifically the boundary on the C_b axis. Figure 4.6 maps the feasible

© University of Pretoria

region of the objective function on the z axis, subject to all the constraints formulated, when focussing on the decision variables of the size A_c on the y axis and C_b on the x axis. With regards to the criterion region in Figure 4.6, because the Pareto optimal set shown in Figure 4.5 (a) is always on the boundary of the feasible criterion region, the weighted sum method provides a necessary condition to always have an optimal solution [50].

The general trend in Figure 4.5 is a strongly Pareto optimal set in which an increase in battery capacity reduces both inconvenience and loss of power supply. To eliminate the loss of power supply during load shedding, households need a minimum 4.8 kWh battery capacity, alongside scheduling. Interestingly, the minimised inconvenience of 100 minutes does not change when the battery capacity changes from 4.8 kWh to 7.2 kWh. This is due to the fixed weighting factors used. The optimal schedule and inconvenience for appliances using battery energy storage is shown in Figure 4.4.

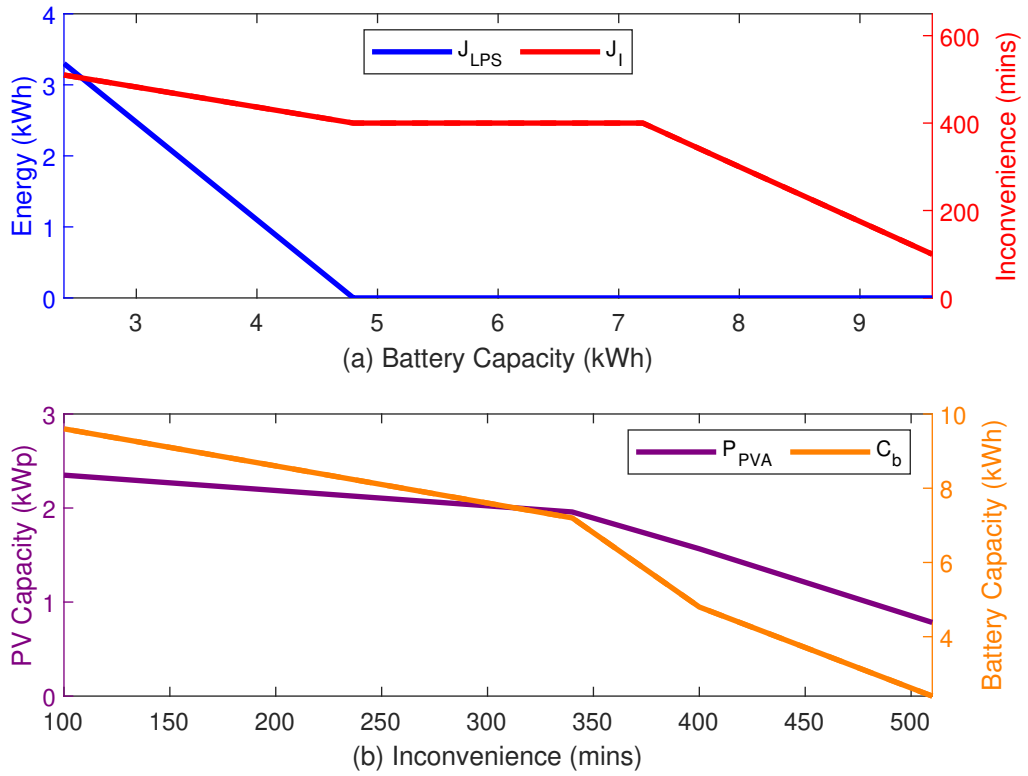

Figure 4.5. (a) The Pareto set of solutions for sizing the batteries when scheduling with load shedding and battery energy storage, and (b) the Pareto set of solutions for sizing the batteries and panels when scheduling with load shedding, panels and batteries using Figure 4.1.

© University of Pretoria

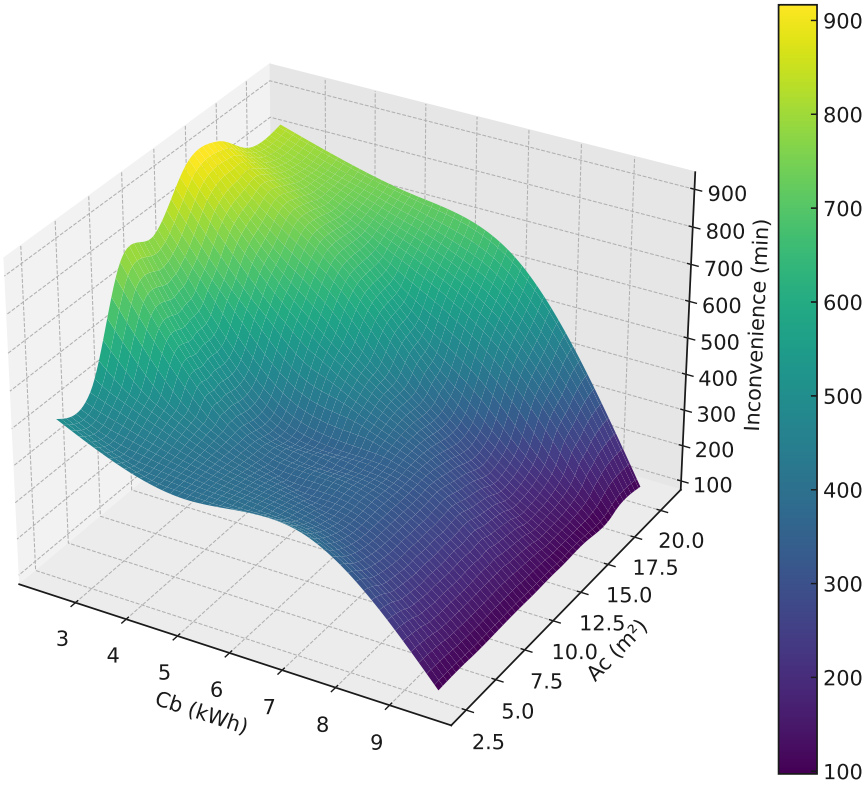

Figure 4.6. The feasible criterion region to determine the sizing decision variables subject to the constraints of scheduling with load shedding using Figure 4.1.

4.4 OPTIMAL STRATEGY FOR SCHEDULING WITH LOAD SHEDDING, A SOLAR PANEL ARRAY AND BATTERY ENERGY STORAGE

The fourth strategy considers HAS and the introduction of both a battery energy storage and a solar panel array. The sizing decision variables A_c and C_b are determined in the optimisation. The LPS for the fourth case study is described as

$$J_{LPS} = \sum_{j=1}^{8} \sum_{k=1}^{N} \left[P_j u_j^{\text{opt}}(k) - P_{gr}(k) u_{LS}(k) - P_b(k) - P_{PV}(k) \right] \Delta t. \tag{4.3}$$

The inconvenience and loss of power supply caused by load shedding, subject to the constraints in Section 3.8, is again minimised in a MOMINLP with $w_1 = 0.5$ and $w_2 = 0.5$. As a result, appliance scheduling with an optimally sized 9.6 kWh battery and 2.4 kWp panel array eliminates loss of power

© University of Pretoria

supply and reduces inconvenience by 92%. The optimal schedule for appliances using batteries and solar panels is shown in Figure 4.7 (a). Unlike HAS+LS+BESS, the addition of solar panels reduces the strain on the grid to charge the battery, as seen in the power flow in Figure 4.7 (b), in which the battery energy storage is charged by the solar panel array. Table 4.4 recommends that households invest in a 2.741 kWp solar panel array and 9.6 kWh battery capacity in 2023, with the scheduling of appliances in Figure 4.7 to minimise load shedding inconvenience and eliminate loss of power supply. The 2.741 kWp panel array translates into approximately five 550 Wp standard panels.

© University of Pretoria

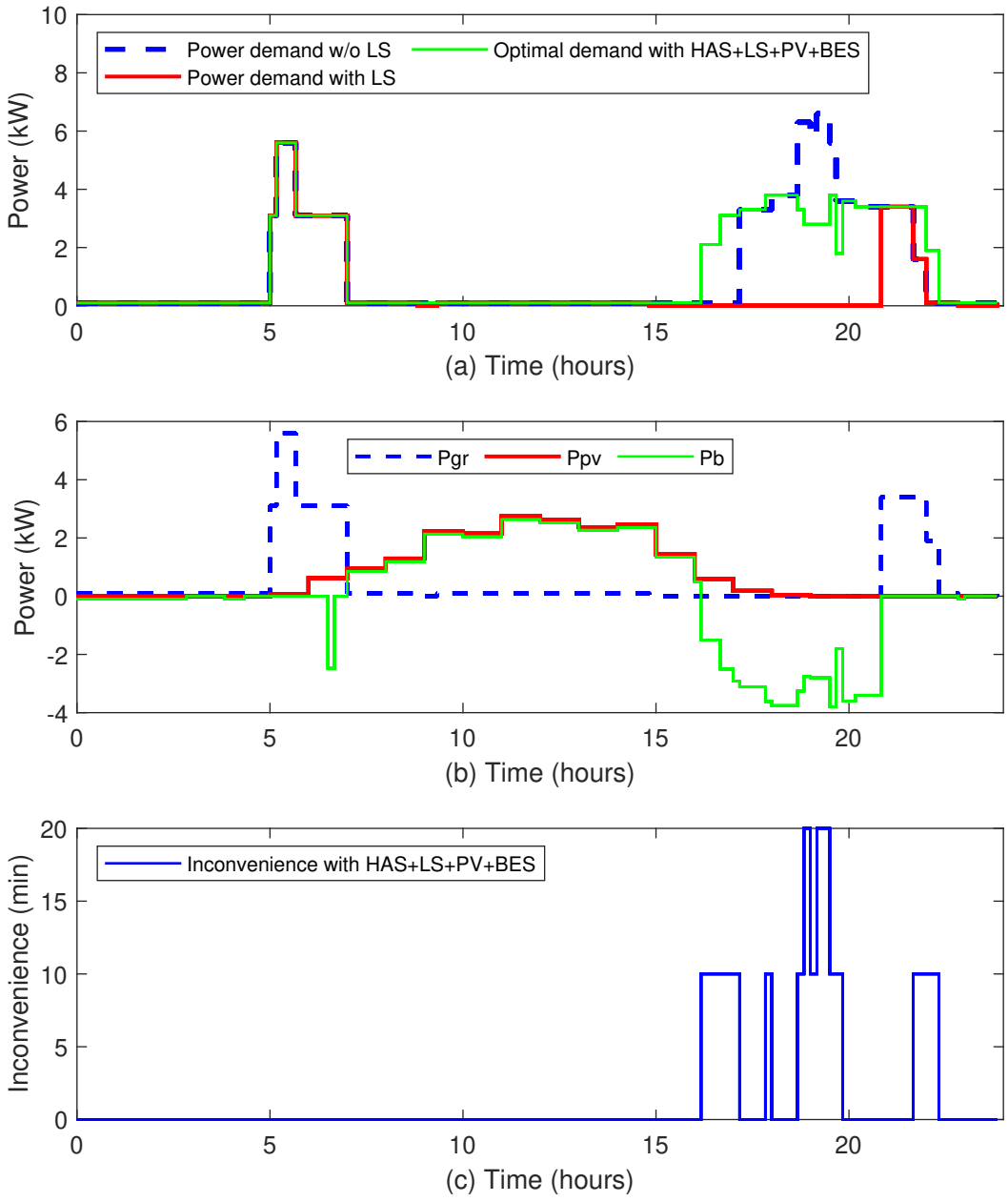

Figure 4.7. (a) The optimal power demand for scheduling with load shedding using Figure 4.1, a solar panel array and battery energy storage, (b) with the optimal power flow management of the energy sources (c) and the inconvenience.

© University of Pretoria

CHAPTER 5 LOAD PROFILE CLUSTERING

The persistent challenge of ensuring reliable and equitable energy access in African countries remains a formidable barrier to socio-economic development. Despite significant progress in recent years, a substantial part of the population remains trapped in energy poverty [51], the International Energy Agency highlighting that in its latest report, an estimated 600 million people in Africa lack access to electricity [52]. Coupled with the challenges of energy poverty, areas on the continent that have electricity only have access to it for limited periods. In the most populous nation on the continent, Nigeria, customers connected to the grid can experience an average of 28 power outages per day [53]. This stark reality not only hampers economic growth but also exacerbates social inequalities, making the search for sustainable energy solutions more critical than ever [54].

Furthermore, many African nations face the widespread problem of load shedding, a strategy used by utilities to prevent total grid collapse by coordinating power outages in response to insufficient generation capacity [55]. Load shedding is used to manage limited energy availability. This strategy, while a temporary solution to a constrained grid, inflicts profound disruptions on daily life and economic activities, underscoring the fragility of the existing energy infrastructure. Recently, the unplanned loss of power generation capacity from coal power plants has caused prolonged and repeated load shedding throughout the country in Egypt [56]. During the load shedding power cuts in Zambia, charcoal is used to cook in 96% of homes in low-income communities, 86% in middle-income communities, and 73% in high-income communities [6]. In Zimbabwe, 60% of the population has experienced food spoilage due to electricity interruptions, which causes energy expenses to account for 16% of their income when there is no power outage. These costs increase to 64% in low-density regions [7]. Although Zambia experiences 6 to 10 hours of load shedding per day [11], Ghana endures up to 16 hours of daily outages [12]. South Africa, in particular, exemplifies this crisis, where load shedding has become a frequent occurrence, reflecting broader challenges in energy policy, generation capacity, and resilience

© University of Pretoria

of infrastructure [10]. The country's struggle to balance energy supply with growing demand highlights the urgent need for innovative scheduling strategies that integrate renewable energy sources. One of the potential solutions could be the use of solar home systems (SHSs) in clustered homes as a strategy to mitigate the impacts of load shedding. Using the abundant solar energy resources available on the continent, SHS offers a scalable and decentralised solution to energy insecurity, particularly in regions plagued by frequent power outages. Therefore, this study explores the optimal integration of SHS's sizing and scheduling strategies in clustered residential households with severe national load shedding conditions.

Given the difficulties caused by load shedding, existing residential scheduling strategies are ineffective when considering the compulsion of load shedding. Studies on the scheduling of residential appliances focus primarily on the use of economic incentives, such as real electricity pricing [57], dynamic pricing [58], and real-time pricing [59], to motivate consumer participation in demand-side scheduling activities. The strategies found in studies on scheduling, such as the strategy of using a feed-in and a time-of-use tariff as incentives to participate in scheduling [18], the strategy of using time-of-use with a sensitivity analysis that highlights the importance of a customer's willingness to pay [19], and the strategy of multi-residential voluntary power sharing for scheduling appliances [20], all are based on two assumptions: 1) the initial assumption is that scheduling is solely driven by time-of-use and other electricity tariffs; 2) the second assumption is that the customer is the exclusive decision maker when it comes to participating in scheduling. Our study does not make these assumptions; it emphasises that load shedding is different from tariff scheduling because scheduling decisions during load shedding are outside the voluntary control of customers. Furthermore, the strategy to size solar panels, wind turbines and batteries in response to load shedding in [22] is limited by the assumption that load shedding outages are always known by the end user, well in advance and with certainty. This approach, while valuable, does not fully encapsulate the dynamic and often low predictability of load shedding, thereby limiting its practical applicability in real-world scenarios where users must adapt to sudden power outages. Therefore, this study addresses a new problem: clustering, scheduling, and sizing under the compulsion of load shedding, seeking an optimal strategy for home appliance scheduling and sizing in response to this problem.

Clustering is introduced to provide broader strategies for different household archetypes to address the unique challenges presented by load shedding. The clustering approach of this study is similar to methods used in other studies, such as the multidimensional reduction of time series data from load

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

profiles in [60], and the evaluation of different clustering algorithms, such as K-means, K-medoids, Hierarchical Agglomerative Clustering, and Density-based Spatial Clustering to extract load profiles as discussed in [61]. Specifically, these clustering techniques consider the total daily energy consumption of each household to determine their proximity to the cluster centroids, which is the scalar clustering method also used in this study. Although [61, 62] has found that the K-medoids clustering algorithm tends to provide different results when determining the optimal number of clusters, the K-means, the K-median, and the K-medoids are part of the same K family, with the latter being a derivative of the K-means [63], which instead of basing the centroids on mean values, bases them on the data points themselves (medoids) as representative cluster representatives. The medoid of a cluster is defined as the specific piece of data that exhibits the lowest average dissimilarity when compared to all other points within the same cluster. However, these data points are still the total daily energy consumption; therefore, we would expect a similar outcome to our K-means clustering. Therefore, this study addresses the critical problem of involuntary scheduling and sizing in residential clusters under severe national load shedding. Our study determines the optimal strategy to integrate the solar panel and battery sizing with the scheduling of clusters that reflect different load profiles of households.

5.1 PROPOSED METHOD

Load profile clusters and load shedding power outages are used as demand and supply parameters, respectively, in optimisation of the solar home system. These two parameters are important for system optimisation because they encapsulate demand and supply dynamics. Through a data-informed understanding of the demand (load profile clusters) and the constrained supply (load shedding power outages), power flow can be optimised to minimise inconvenience and the loss of power supply. This approach formulates effective strategies to mitigate load shedding effects, improving the scheduling, sizing, and energy management of home appliances with renewable generation, thus leading to an optimised use of the solar home system. Using suitable clustering techniques, broader strategies are developed to mitigate the negative impacts of load shedding between diverse residential groups. Consequently, this study optimises home appliance scheduling in the midst of uncertain load shedding outages to minimise inconvenience and loss of power supply (LPS) in clustered households. The power supply to the grid is represented as $u_{LS}(k)P_{gr}(k)$ and depends on the load shedding model u_{LS}(k). The residential load is represented as $P_m(k)$ and depends on the cluster m of the K-means load profile clustering algorithm used in this study. The power output of the photovoltaic array is represented as $P_{PVA}(k)$ and the power flow to and from the battery bank is represented as $P_b(k)$.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Among severe load shedding outages caused by unplanned thermal generation losses from ageing power stations, this study uses the novel forecasting model developed in [64] that learns outage patterns from historical data to provide an hourly forecast of load shedding for a year ahead, as illustrated in red blocks in Figure 5.1. Step 1 involves pre-processing the Eskom hourly dataset from 2014 to 2023 by (1) validating that Dispatchable Generation = Residual Demand + Interruption of Supply + Interruptible Load Shed + Manual Load Reduction, and (2) validating that the measured megawatts of load shedding are correctly correlated with the recorded stages of load shedding. If the measured megawatts of load shedding are greater than zero but less than 1 000 MW, then this should be Stage 1 of load shedding. If the measured megawatts of load shedding are greater than or equal to 1 000 MW but less than 2 000 MW, then this should be Stage 2 of load shedding. If the measured megawatts of load shedding are greater than or equal to 2 000 MW but less than 3 000 MW, then this should be Stage 3 of load shedding. If the measured megawatts of load shedding are greater than or equal to 3 000 MW but less than 4 000 MW, then this should be Stage 4 of load shedding. If the measured megawatts of load shedding are greater than or equal to 4 000 MW but less than 5 000 MW, then this should be Stage 5 of load shedding. If the measured megawatts of load shedding are greater than or equal to 5 000 MW but less than 6 000 MW, then this should be Stage 6 of load shedding.

The output of the pre-processing step consists of a dataset with 78 880 rows, where each row represents hourly observations. This dataset contains 35 columns, each column representing a different feature. This pre-processed dataset is then used as the input for Step 2, which uses the mRMR (Minimum Redundancy Maximum Relevance) method to minimise redundant inputs while maximising relevant inputs. The mRMR method produces four time-dependent outputs: Year, Month, Day, and Hour, which are considered optimal features of any time instance denoted by k. These four optimal features are used as input in Step 3 of the K-Nearest Neighbours (KNN) load shedding time series classification forecasting model for load shedding stages $x_{LS}(k)$, based on the previous study [64].

At time k, the load shedding stages are represented as $x_{LS}(k)$, an integer indicator of the severity of power outages $u_{LS}(k)$. The K-Nearest Neighbours (KNN) is a non-parametric method employed in machine learning for various purposes such as classification, regression, and prediction [43]. In this study, it is applied to classify and predict future load shedding stages. The output consists of the load shedding stage and a binary value indicating the presence of a power cut at a particular upcoming time instance k. The KNN algorithm operates on the principle that the closest neighbours to an unknown load shedding pattern at time k offer crucial classification details. A novel KNN

© University of Pretoria

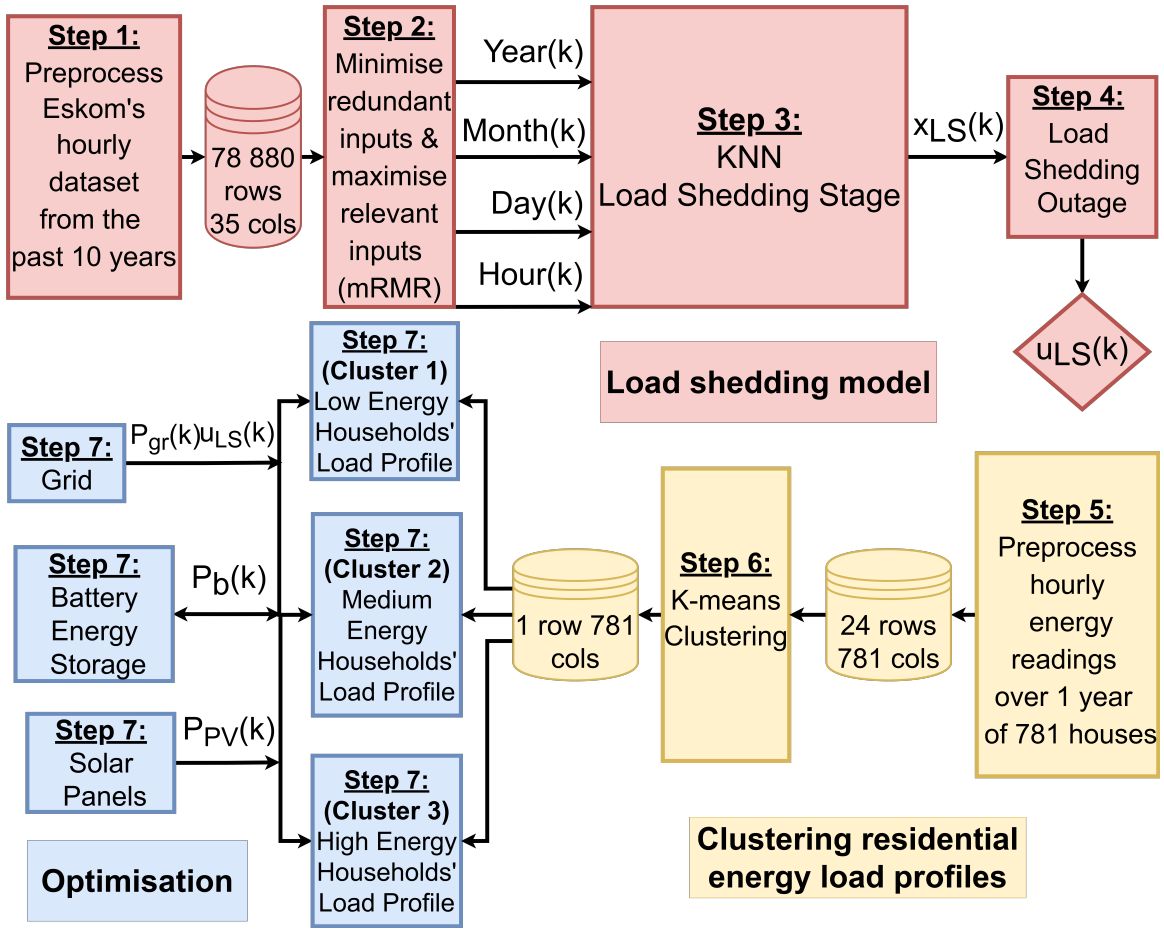

Figure 5.1. Illustration schematic for proposed residential power flow management with renewable energy integration considering load shedding in clustered residential homes.

load shedding algorithm is developed to maximise the classification of a time instance k, originally formulated in the previous study with a 94.4% accuracy [64], to the most common stage of its nearest neighbours,

$$x_{LS}(k) = \underset{x_{LS} \in C}{\operatorname{argmax}} \sum_{n=1}^{K_1} \delta(x_{LS}(k), x_{LS}(y_n)), \tag{5.1}$$

where $x_{LS}(k)$ is the load shedding stage predicted at time k and is an integer variable that can only take values from class C, which is the set of integer class labels. Each value of $x_{LS}(k)$ corresponds to the amount of gigawatts of load demand that must be shed from the grid, varying from 1 000 MW of load demand that is involuntarily shed from end users when $x_{LS}(k) = 1$, to 6 000 MW that is shed when $x_{LS}(k) = 6$. Interestingly, dispatchable generation is always equal to the sum of residual demand and the load that is shed from the grid, and this ensures that grid supply and demand are balanced. K_1 is the number of nearest neighbours at time k, y_n are the closest neighbours at time k, both variables being used to determine how many of the closest data points are considered when making the prediction,

© University of Pretoria

and $\delta(x_{LS}, x_{LS}(y_n)) = 1$ if $x_{LS} = x_{LS}(y_n)$ and $\delta(x_{LS}, x_{LS}(y_n)) = 0$ otherwise. $\delta(x_{LS}, x_{LS}(y_n))$ is a binary function that is essential for counting the occurrences of each stage among the nearest neighbours. In this study, (5.1) determines the most common load shedding stage among the closest neighbours K_1 of a given time instance k in the dataset, and the class that occurs the most frequently among these neighbours is predicted as the load shedding stage $x_{LS}(k)$. This method is characteristic of KNN algorithms, in which the prediction is based on the majority decision of the closest neighbours. A unitary binary power supply represents a continuous grid supply without outages, allowing the grid to meet the demand for power.

$$u_{LS}(k) = \begin{cases} 1, & \text{if } x_{LS}(k) = 0, \text{ no power outage} \\ 0, & \text{if } x_{LS}(k) > 0, \text{ power outage} \end{cases} \tag{5.2}$$

where $u_{LS}(k)$ formulates the load shedding power outage state at time k. This novel time series classification model enables the determination of both the times when load shedding occurs and the duration of load shedding outages.

Load shedding outages have a direct impact on the behavioural patterns of electricity consumption in millions of households, yet these patterns have not been considered in the clustering of load profiles. In Step 5, we preprocess hourly energy readings from 781 households in 2014, the year that marked the start of a decade of severe national load shedding in South Africa, although there has been load shedding since 2008. The preprocessing in Step 5 involves (1) retaining zero consumption values and (2) identifying valid daily energy consumption load profiles. Retention of zero values from the original meter readings of households in countries that experience load shedding is important for several reasons, including the following.

1. Retaining Meaningful Data: Zeros may represent significant events, such as load shedding outages in electricity consumption datasets. Removing or forward-filling these zeros could lead to loss of critical information and misinterpretation of the data.
2. Accurate Representation: In the context of electricity consumption, zero readings can accurately indicate periods of no usage due to outages. These data observations are essential for understanding and clustering behavioural patterns, especially under specific conditions such as load shedding.
3. Improved Clustering Results: Including zeros ensures that the clustering algorithms consider all relevant variations and patterns in the data. This leads to more accurate and meaningful

© University of Pretoria

clustering results, as the algorithm can account for periods of zero consumption.

The output of Step 5 is a preprocessed dataset of 24 rows, where each row represents the yearly average hourly metered readings of household energy consumption, and 781 columns where each column represents a different household. The 24 rows and 781 columns are the dimensional input to the K-means clustering in Step 6, the output is a single row with 781 columns containing each household's cluster number from 1 to K.

K-means clustering aims to minimise the variations within the clusters by iteratively updating the means of the initial randomly selected cluster centroids, ensuring that intra-cluster variations are low, while inter-cluster variations are high. Among several K-means algorithms with different loss functions such as using the mean squared error reconstruction loss [65], K-means clustering using [66]'s least squares quantisation in pulse-code modulation (PCM) loss function, initially used in the PCM sampling theorem, is better suited for load profile clustering in this study. The use of K-means clustering remains popular in the fields of data mining, computer vision, pattern recognition, and quantum computing [67]. The application of clustering algorithms to achieve optimal load profiling and load shapes derived from electricity metered data, in identifying the most suitable consumption behaviours, remains an essential field in demand response applications, and clustering under severe national load shedding is a new field that has not been explored in load profile clustering. Thus, this study makes a contribution to this new field of clustering of load profiles under severe national load shedding. This study not only considers realistic load shedding outage patterns that are either oversimplified or not considered at all in demand side management but also considers load shedding in the context of load profile clustering.

5.2 K-MEANS CLUSTERING DISTANCE SELECTION

The problem of calculating pairwise Euclidean distances with K-means algorithms based on the widely used PCM loss function [66] arises when dealing with high-dimensional datasets [68]. The Euclidean distance is a commonly used measure of similarity between two data points in a multidimensional space. It is calculated as the square root of the sum of squared differences between the corresponding elements of the two data points. In the K-means algorithm, during the assignment operation, each data point is assigned to its nearest cluster centroid by calculating the Euclidean distance between the data point and all cluster centroids. This process is repeated for every data point in the dataset until convergence [69]. However, in high-dimensional spaces, the number of dimensions (features) can be extremely large, which leads to a significant increase in the computational complexity of calculating

© University of Pretoria

Euclidean distances. As the dimensionality of the data increases, it becomes increasingly sparse, and the distances between the data points become less meaningful as the number of dimensions increases. This means that the Euclidean distance between two data points in high-dimensional space may not accurately reflect their true similarity or dissimilarity. Moreover, computing pairwise Euclidean distances in high-dimensional spaces requires a large number of arithmetic operations, which can be computationally expensive and time consuming. This can result in inadequate efficiency and scalability issues when dealing with large datasets or real-time applications.

The study in [70] addresses this high-dimensionality problem by pruning unnecessary distance computations with their proposed lower bound-based K-means clustering (LBKC) algorithm. However, their LBKF framework is defined and illustrated only for 2-dimensional clustering, which does not address high-dimensional clustering. Various techniques have been proposed to reduce the dimensionality of the data while preserving its structure. [71] proposes a feature selection and removal technique for dimensionality reduction that involves two steps: relevant feature selection and redundant attribute removal. The process identifies the most relevant attributes for clustering by applying similarity measures such as simple matching coefficient (SMC), Jaccard coefficient (JC), and frequency probability-based similarity measure (FPSM) to each dataset. An empirically chosen threshold value (0.80 to 0.96) determines attribute relevance, balancing informative attributes, and dimensionality reduction. However, [71] does not provide a comparison between the number of reduced features and the performance of the K-means algorithm, in terms of reduced computations or improved computation speed. Without these comparisons, the effect of the feature-selection-attribute-removal algorithm on the high-dimensional K-means problem cannot be quantified. In addition, [72] introduces a Transformer Autoencoder for K-means Efficient clustering (TAKE) to address high dimensionality. The network has two components: the Transformer AutoEncoder (TAE) for feature reduction and the KNet for clustering. The encoder maps the high-dimensional data into a low-dimensional feature representation and the decoder reverses it. By minimising an appropriately defined reconstruction loss, the TAE is able to derive a meaningful representation. In the context of autoencoders, reconstruction loss quantifies the dissimilarity between the original high-dimensional input data and the reconstructed low-dimensional output produced by the autoencoder. The loss quantifies how well the autoencoder can reconstruct the input data from its compressed representation. However, [72] does not consider the common criticism of using autoencoders that minimise reconstruction losses, which is their dependence on labelled data. This limitation confines their use to supervised machine learning tasks such as classification, rather than unsupervised applications such as clustering.

© University of Pretoria

Principal component analysis (PCA) is another approach commonly used in the reduction of high dimensions in clustering [73]. PCA finds dimensions that display the highest data variance, which is effective when the clusters are distinguishable and linearly separable along the high-variance axes [74]. However, PCA assumes a linear structure in the data, which may not always be the case. In fact, we cannot assume a linear structure in the metered electricity consumption of several households over time. The electricity consumption patterns of households are influenced by various factors such as daily routines, weather conditions, occupancy, and individual preferences. These factors can result in nonlinear relationships and complex patterns in electricity consumption. For example, electricity consumption can vary throughout the day, with a higher usage during peak hours and a lower usage during off-peak hours. In addition, individual households may have abnormal electricity load profiles in response to disturbance-induced short-term and long-term load reductions [75], as well as varying consumption patterns driven by their unique needs and behaviours in reaction to mandatory power outages due to load shedding, the timing and duration of which have low linear predictability [64]. Taking these factors into account, it is more appropriate to assume that the electricity consumption of households over time follows a nonlinear structure rather than a linear one. Therefore, when reducing and clustering metered electricity consumption data, it is important to consider nonlinear techniques and algorithms that can capture the complex patterns and relationships in the data.

In this study, to solve the problem of high dimensionality associated with the computation of pairwise distances using the commonly used K-means clustering Euclidean distance metric presented in [66], we compare four different distance metrics, namely the squared Euclidean distance [76, 77], the Manhattan or City Block (CB) distance [78], the Cosine distance [79] and the correlation distance [80], to systematically determine the most suitable distance cost function for high-dimensional metered electricity data on household energy consumption. The determination of a suitable distance metric is essential for K-means algorithms to form well-separated clusters of data observations [81], because distance metrics have different characteristics that can lead to unsuitable cluster formations in the clustering of load profiles. Therefore, it is important to carefully determine the most suitable distance metric that aligns with the nature of the data on household energy consumption and the objectives of the clustering task. The Manhattan or CB distance used in [78, 82] is named after the grid-like layout of Manhattan streets, where the distance between two points is the sum of the absolute differences of their coordinates. In load profile clustering, this grid-like nature refers to the structure of data points in a time-series energy dataset, specifically the two-dimensional structure of time and load. The time dimension represents chronological order, and the load dimension represents energy consumption at

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

each timestamp. This creates a grid-like structure where each data point is a coordinate on a grid. However, because time-series data can also have additional dimensions, such as weather conditions and temporal characteristics of occupant behaviour driven by their heating/cooling demand [83], the suitability of the Manhattan metric must be determined with caution. Compared with the Euclidean distance, the Manhattan distance is less sensitive to outliers because it sums absolute differences and, in high-dimensional data, outliers disproportionately affect cluster centroids, leading to skewed-distributions of clusters.

The Cosine and the Squared Euclidean distances are criticised for their inability to deal with transformations on the time axis [84]. This criticism points out that these distance metrics compare the corresponding points directly without accounting for possible shifts or distortions in the time series. For example, if there are two time series representing the timestamp dimensions of hourly electricity consumption for two different households in a dataset, in which the first time series (first household) peaks on Monday morning and evening, plateaus on Tuesday without a significant change in consumption, dips on Wednesday afternoon, before it plateaus on Thursday and peaks again on Friday evening. The second time series (second household) follows a similar consumption pattern, but shifted one day later. The Cosine and Squared Euclidean distance metrics would have large distance values because the corresponding timestamp dimensions of the two households do not match. Therefore, if used in K-means which clusters based on the minimum distance between time series, it would erroneously miscluster these two similar energy consumption patterns as dissimilar. A reasonable explanation for why there would be temporal shifts in time series data is the cyclic and periodic nature of weather events [85]. For example, weather conditions starting in one area and altering energy use in houses, then moving to another area, would cause delayed consumption patterns in the different time series.

The scale independence of the cosine distance measure in [86] and the Pearson correlation distance measure in [87] is a fundamental characteristic that differentiates the angle and correlation-based metrics from other distance measures in the clustering of high-dimensional load profiles. Notwithstanding their extensive use, Minkowski-based distance measures exhibit scale variance. Consequently, distances calculated using metrics such as Squared Euclidean may be distorted contingent upon the dimensional units, particularly if the dataset is not subjected to normalisation prior to the implementation of clustering algorithms. When calculating the cosine distance between two vectors, the magnitude of the vectors does not influence the result. The cosine measure normalises the vectors by dividing the dot

© University of Pretoria

product of the vectors by the product of their magnitudes. The correlation measure is normalised by dividing the covariance of the variables by the product of their standard deviation. This normalisation ensures that the cosine and correlation distance is not affected by the scale of the vectors. In multi-featured load profile clustering where diverse features with different units and high-dimensional time variance (such as temperature, humidity, and energy consumption) need to be compared, the cosine and correlation measures enable a meaningful assessment of similarity or dissimilarity without being influenced by the specific magnitudes of the features. However, in developed countries, the energy consumption load patterns are generally more stable and predictable, which leads to clustering results that are within two standard deviations of each cluster, respectively [88]. Normalisation techniques can effectively smooth out minor variations and highlight overall trends and archetypes, making it easier to analyse and interpret the data. When normalisation is not considered by the appropriate selection of the distance measure, the electrical consumption data set of residential households is often normalised with the deminning normalisation technique of subtracting the minimum energy use for that home in all its dimensions and dividing by the difference between its maximum and minimum energy use [89]. Normalisation of energy consumption datasets is effective in the developed world because of stable and predictable energy consumption patterns, well-established infrastructure, and consistent supply, which allow for the smoothing out of minor variations to highlight overall trends. However, this approach is ineffective in capturing complex and unique energy consumption patterns in the developing world, where significant variability arises from irregular supply, varying access levels, and socioeconomic disparities. Zero value outliers in these regions are indicative of energy poverty, such as complete power outages and severe national load shedding. Filtering out these outliers through normalisation can lead to a misrepresentation of the true energy load profiles in developing countries. Although a similar observation is made in [90] about zero consumption values and outliers in a dataset with low-income and rural households, the study uses a different normalisation technique (unit norm and zero-one instead of deminning), rather than systematically determining an appropriate high-dimensional distance metric that incorporates zero consumption values and outliers in K-means clustering. In addition, the research in [90] also presents the concept of pre-binning, which involves clustering data that has already been grouped. This approach overlooks the crucial task of determining the optimal number of clusters and raises two critical research questions: 1) Is pre-binning an effective grouping method, and if so, 2) Why is it necessary to perform clustering on pre-grouped data?

Therefore, this study develops a novel Manhattan distance metric d to mitigate the curse of dimensionality for clustering of residential load profiles in developing countries with severe national load

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

shedding which is formulated as

$$d(e,c) = \min \sum_{m=1}^{K_2} \sum_{p=1}^{781} \sum_{k=1}^{N} |e(p,k) - c(m,k)| \, \Delta t, \tag{5.3}$$

where p signifies the index of the 781 households at a given time instance k, with $N = 144$ representing the total number of sampling periods. $\Delta t = 10$ minutes delineates the duration of the sampling and scheduling interval. The parameter m corresponds to the cluster index, with K_2 denoting the upper limit of the clusters obtained by the K-means clustering. The variable e characterises the energy consumption of residential households from a dataset encompassing zero values due to load shedding and outliers indicative of the energy disparity among households in developing nations, as a constant vector input. The inclusion of these outliers, as discussed in [91], is necessary for an accurate analysis of energy consumption patterns, as they represent significant energy consumers and hold the highest potential for energy efficiency interventions. The parameter c represents a vector decision variable of the centroids of the clusters, which must be determined iteratively to find the optimal distance centre for each cluster, and the centroid $c(m,k)$ is the component-wise median of the points in that cluster. Using this metric, each term $|e(p,k) - c(m,k)|\Delta t$ measures the absolute variation between the actual energy consumption and the centroid of the cluster at each time instance scaled by the duration of the interval. Minimising the sum of the variances in daily energy consumption to the cluster centroids in all households, sampling periods, and clusters provides a comprehensive measure to maximise total inter-cluster dissimilarity and minimise intra-cluster similarity, enabling an effective clustering approach that accounts for the unique load shedding outages and energy consumption patterns in developing countries.

Although using dynamic time warping (DTW) enables the alignment of different household time series by allowing elastic shifts on the time axis, making it an effective approach to compare high-dimensional time series with similar but misaligned patterns [92], a simpler approach is used in this study to reduce dimensionality while highlighting the most relevant patterns of electricity consumption for clustering purposes. The main argument against using DTW for load profile clustering is that it can lead to misleading clusters by ignoring important temporal shifts. For instance, one time series might have peaks in the early morning and evening (household load profile), while another has peaks in the late morning and afternoon (business load profile). DTW would cluster these two time series erroneously. In addition, considering that the applications of load profile clustering are more concerned about the energy consumption patterns between households, often between millions of households, and less concerned about the seasonal consumption patterns within a small number of households, the need for

© University of Pretoria

temporal alignment is given less weight than the need to reduce the energy dataset dimensions. It also remains unclear whether the Silhouette and Davies-Bouldin indices show significantly different results when comparing K-means and K-medoids with a Euclidean distance versus dynamic time warping and limited dynamic time warping distances [93]. The dataset in this study has the same structure as most time series data for hourly energy consumption. However, it presents a different number of metered houses and their unique energy consumption patterns in an environment of national load shedding. Within the structure of time-series data that capture hourly energy consumption across 781 households, each row corresponds to an hourly timestamp over the span of a year (8 760 rows), while each column denotes a distinct household (781 columns). Consequently, each hour creates a separate dimension. As the dimensions increase, the Euclidean distance between any pair of data points diminishes in significance. In an 8 760-dimensional space, the data points tend to be equidistant from each other, which impedes the K-means algorithm from determining discernible clusters.

To create discernible clusters with non-equidistant data points, this study reduces the 8 760 dimensions to 24 by averaging the daily energy consumption for each of the 781 different households. The 24 dimensions and 781 households (features) with the optimal number of clusters K are input to the clustering of K-means in MATLAB. The average total daily energy consumption further reduces the 24 dimensions to 1, creating a 1-D K-means clustering approach for load profiles. The clustering approach to investigate daily electricity consumption patterns (DECP) using K-means clustering is not unique to this study. For example, the study in [94] investigates DECP, however, because the study uses normalisation to make the load profiles of different households comparable, allowing the clustering algorithm to focus on the shape of the consumption, normalisation ignores the absolute levels of energy consumption. This leads to the loss of crucial information for energy management and resource allocation, planning, and integration, based on the total energy demand of different households. The study in [95] also investigates the DECP; however, it replaces zero consumption values with regression values. This assumes that the only reason there would be no consumption is due to an invalid meter reading; this study investigates this assumption. In addition, the study in [96] investigates both the DECP and the daily gas consumption patterns (DGCP); however, it addresses a research question that has been extensively studied in the clustering literature, which investigates the variability in peak demand time in different customer segments and its effects on daily electricity consumption patterns; this is evident from the different clusters of the study, which include the “Midday peak”, “Late afternoon” and “Evening” archetypes.

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

5.3 K-MEANS CLUSTERING OF LOAD PROFILES CONSIDERING LOAD SHEDDING

Figure 5.2 illustrates the proposed method for segmenting the energy patterns of households by clustering using K-means to establish representative energy consumption archetypes of different households. Step 1 determines the total daily energy consumption of the 781 households as grey dots on a scalar axis with 0 kWh as the lowest and 40 kWh as the highest daily demand.

There is an initialisation problem inherent in the K-means algorithm when the initial cluster centroids are selected in a stochastic manner, causing the iterative process of assignment and refinement to continue until the centroids converge to a local minimum far from the global optimum. This stochastic initialisation often leads to suboptimal clustering outcomes [97]. To address this problem, the *k*-means++ algorithm improves the initialisation by selecting the first centroid stochastically, and each subsequent centroid is chosen with a probability proportional to the squared distance from the closest already chosen centroid. This approach spreads out the centroids, leading to better clustering results and faster convergence to the global optimum [98]. The *k*-means++ initialisation is inherently sequential, requiring multiple passes over the data, which limits its scalability, especially for large datasets. The scalable *k*-means++ algorithm in [99] addresses this by parallelising the initialisation process. Instead of selecting centroids one by one, it selects multiple centroids in parallel using an oversampling approach and then reclusters these centroids to form the final K centroids. This method drastically reduces the number of passes over the data, making it suitable for large-scale data clustering. As shown in Step 2 of Figure 5.2, the centroids are determined using the *k*-means++ algorithm with a parallelised initialisation process in MATLAB.

The optimal number of clusters in these data points is determined by K, and if $K = 3$ then the three clusters represent the daily energy consumption patterns of the 781 households. The centroid, representing the centre of a cluster, is initially determined using the *k*-means++ algorithm and then recalculated as the mean of all data points in the cluster. The distances to cluster centroids/mean represent the variances of each cluster. The K-means clustering algorithm minimises the sum of distances to the mean, essentially minimising the variances for each cluster.

If the minimum variance from the mean of the cluster for each household's daily energy consumption is zero, then the household's daily energy consumption is the same as the average daily energy consumption for its cluster. If the minimum sum of distances to the mean for every household was

© University of Pretoria

zero, then variances would be zero and each household's daily energy consumption would be identical to the average daily consumption of its cluster. The minimum variance is shown in the energy variance graph for each household in Figure 5.2, where the blue represents the variances of the daily energy consumption from the cluster 1 mean, the red represents variances from the cluster 2 mean and the green represents the variances from the cluster 3 mean.

Each household is clustered based on its lowest variance or minimum distance to either cluster 1, cluster 2 or cluster 3's mean. For example, in Figure 5.2, the lowest variance for households 1 to 338 is the minimised distance to the mean of cluster 1 in blue; therefore, they are grouped into cluster 1. The lowest variance for households 339 to 666 is the minimised distance to the mean of cluster 2 in red; therefore, they are grouped in cluster 2. The lowest variance for households 667 to 781 is the minimised distance to the mean of cluster 3 in green; therefore, they are grouped in cluster 3.

The average minimum variance for all households in cluster 1 is 2.32 kWh and the minimum sum of distances to the cluster 1 mean is 784.81 kWh. The average minimum variance for all households in cluster 2 is 3.07 kWh and the minimum sum of distances to the cluster 2 mean is 1 006.28 kWh. The average minimum variance for all households in cluster 3 is 6.87 kWh and the minimum sum of distances to the cluster 3 mean is 790 kWh.

First, we prepare the original hourly energy data set [100]. Second, we preprocess data from the hourly energy consumption profiles of 781 households over a metering period of one year; Data cleaning includes removing missing values and converting the amperes to power by multiplying with a conversion factor with an assumed 230 V constant voltage. This approach is similar to the data processing strategy used in [101] and [102], which determine the energy used for each hour, measured in kilowatt hours, assuming constant voltage. Third, we determine the optimal number of clusters. Fourth, we use K-means clustering to determine each of the K_2 clusters. Finally, we determine the K_2 clusters that represent the household load profiles for low, medium, and high energy consumption.

5.4 OPTIMISATION FORMULATION

The clustering of residential load profiles in the yellow blocks of Figure 5.1 through Steps 5 and 6 results in three load profile archetypes in Step 7. Cluster 1 represents low-energy consumption households, cluster 2 represents medium-energy consumption households, and cluster 3 represents high-energy

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

consumption households. According to [103], 68.1% of the respondents consider alternative energy sources due to load shedding. However, 70.8% indicate that these alternatives are too expensive, which discourages the use of solar home systems. Optimisation is introduced to address the constraints faced by different clusters, offering optimal sizing and scheduling strategies to mitigate the impact of load shedding.

The operation and capacity of the hybrid solar and battery home system for the clustered residential load profiles for households with high, medium and low energy consumption under severe national load shedding is formulated in this section. The photovoltaic system consists of a combination of several solar cells that convert solar energy into direct current power. The maximum hourly output power (kW) of the PV system, $P_{PVA}(k)$ is formulated in [64] as

$$P_{PVA}(k) = \eta_{PV} I_{PV}(k) A_c, \tag{5.4}$$

where η_{PV} is the efficiency of the solar PV generator, $I_{PV}(k)$ is the solar irradiation (kWh/m^2) incident on the PV panels in the kth time interval, and A_c is the surface area of the given solar panels (m^2). In this study, A_c is a sizing decision variable determined in the optimisation algorithm. The lower and upper bounds of the cross-sectional surface area sizing decision variable are subject to the available roof space of the house. The battery energy storage system is formulated as

$$SOE_b(k) = SOE_b(0) + \eta \sum_{\tau=0}^{k-1} \text{sgn}\left[P_b(\tau)\right] \Delta t, \tag{5.5}$$

$$\eta = \begin{cases} \eta_C, & \text{if } P_b(\tau) > 0 \\ \eta_D, & \text{if } P_b(\tau) < 0 \end{cases} \tag{5.6}$$

where P_b represents the power of the battery, n_C denotes the efficiency of the converter during charging, n_D stands for the efficiency of the converter during discharging, and SOE_b indicates the state-of-energy (SOE) of the battery, which is measured in kilowatt-hours (kWh). The initial SOE is denoted as $SOE_b(0)$ and is assumed to be 50% of the battery capacity C_b. The operational constraints of the battery energy storage system should include the lower and upper boundaries on the battery,

$$SOE_b^{min} \leq SOE_b(k) \leq C_b, \tag{5.7}$$

which ensure that the battery remains within the specified limits of its minimum capacity SOE_b^{min} and maximum rating during both discharging and charging. The sizing decision variable C_b is used to size the battery in the optimisation algorithm. The lowest SOE of the battery is restricted by the recommended depth of discharge (DOD) to prevent battery degradation that may occur if it were fully

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

discharged,

$$SOE_b^{min} = (1 - DOD)C_b. \tag{5.8}$$

The operation and capacity of the solar home system is formulated as an optimisation problem. The first objective is to minimise the LPS represented as

$$J_{LPS} = \sum_{m=1}^{K_2} \sum_{k=1}^{N} \left[P_m u_m^{\text{opt}}(k) - P_{gr}(k) u_{LS}(k) - P_b(k) - P_{PVA}(k)\right]^2 \Delta t, \tag{5.9}$$

where $u_m^{opt}(k)$ represents the optimised binary switching status of cluster m at time k, P_m is the baseline power of the mth cluster and the objective J_{LPS} is the LPS due to load shedding. The second objective is to minimise the inconvenience of the load shedding formulated in [64],

$$J_I = \sum_{m=1}^{K_2} \sum_{k=1}^{N} \left[u_m^{bl}(k) u_{LS}(k) - u_m^{opt}(k)\right]^2 \Delta t, \tag{5.10}$$

where $u_m^{bl}(k)$ represents the baseline operating status of the mth cluster.

$$J = w_1 J_{LPS} + w_2 J_I \tag{5.11}$$

where w_1 and w_2 denote weights, with inconvenience measuring time and LPS measuring energy. The total sum of the weights is 1. Both the previous study [64] and this study are the first to quantify the time and energy lost as a measurable impact of load shedding. The inconvenience objective focuses on maintaining the regular schedule even during a LPS, while the LPS objective is concerned with ensuring that each appliance is scheduled, even if it results in inconvenient rescheduling. J aims to find the optimal balance between adhering to the baseline schedule and ensuring that residents have maximum access to electricity.

© University of Pretoria

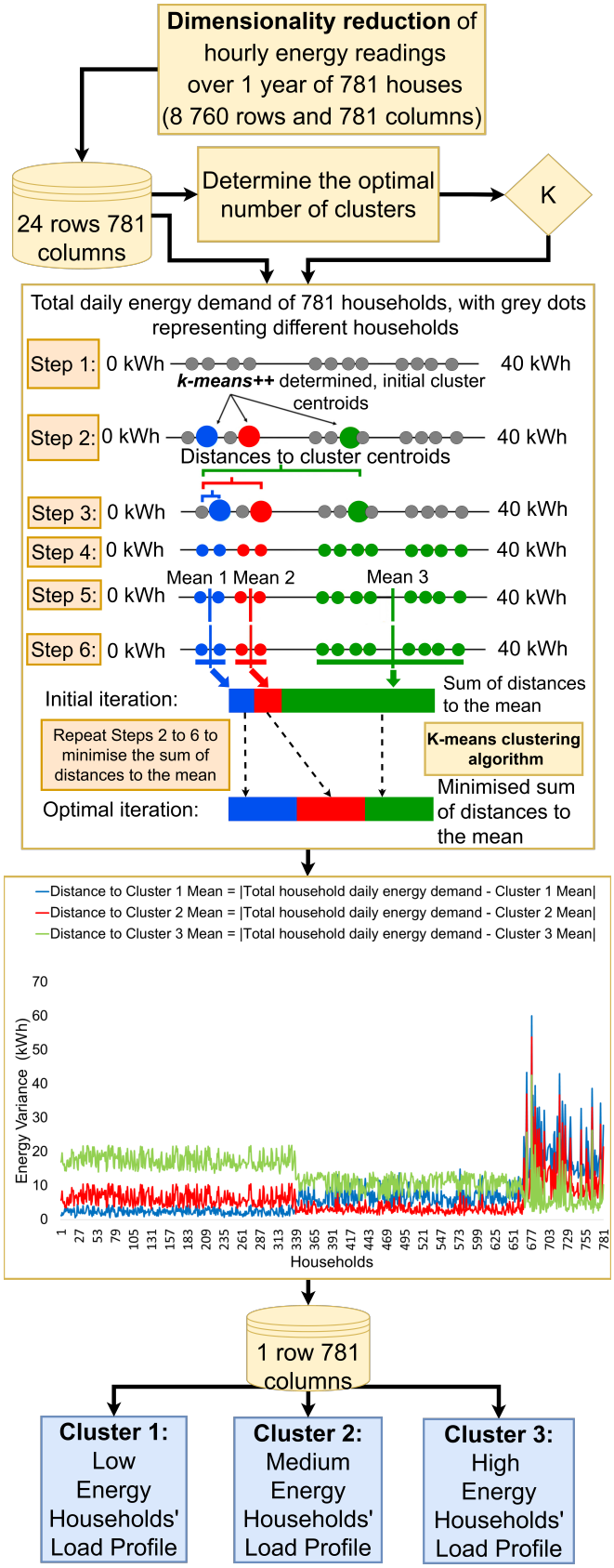

Figure 5.2. An illustration of the proposed household energy consumption clustering using K-means.

© University of Pretoria

CHAPTER 6 CASE STUDY: CLUSTERING RESIDENTIAL HOUSEHOLD LOAD PROFILES

In this second case study, two sets of data are used: the first consists of data on load shedding power outages from the past 10 years that cause rolling blackouts in residential areas, while the second data comprises electricity meter readings from 781 households in South Africa in 2014, collected every 5 minutes. This approach with two aspects allows us to propose a solution that not only forecasts supply-side power disruptions, such as load shedding, but also identifies demand-side clusters within energy consumption profiles for households, enabling a dual-approach energy management strategy to mitigate the impact of load shedding power outages across diverse energy needs of different households.

6.1 OVERVIEW OF THE METERED HOUSEHOLDS

An analysis of the South African Domestic Electrical Load (DEL) datasets in [91], used as the second data stream for this case study, sheds light on the intricate interaction between socioeconomic factors, technological advances, and environmental considerations in the context of energy consumption and management within residential sectors. These datasets, which cover household electricity consumption and sociodemographic survey data in various environments, climatic zones, income groups, electrification stages, and dwelling structures in South Africa, serve as an important resource to manage energy demand. However, the research question of how to determine a representative baseline load profile that is inclusive remains elusive. The types of data included in the DEL datasets are as follows:

1. Metered household electricity consumption data: this includes data on electrical current, collected at 5-minute intervals aggregated over an hour.
2. Sociodemographic survey data: this data captures socioeconomic characteristics of metered households in South Africa.

© University of Pretoria

3. Additionally, there is a harmonisation of the survey data for frequently used variables, such as the appliance variables, known as the DEL Survey - Key Variables dataset.

The DEL datasets are the largest and longest collection of residential electricity consumption in South Africa [90]. This study only considers electricity metering data from the Domestic Electrical Load Metering, Hourly (DELMH) dataset [100] and appliance data from the Domestic Electrical Load Survey (DELS) dataset [104] using profileID and AnswerID found in the Domestic Electrical Load Survey - Key Variables (DELSKV) dataset [105]. The broader survey data in [104], which includes household and dwelling characteristics such as demographics, education, employment, income, income sources, number of household occupants with income, ownership of the house, gender of the head of the house, house roof material, floor area, rooms and water source, will be considered for future multivariate clustering research on households with severe national load shedding.

The metered readings of the DELMH dataset consist of hourly measured time-series current values in amperes for the most recent year in the dataset, 2014. The equivalent energy consumption, quantified in kilowatt hours, is calculated by presuming a fixed voltage of 230 V (the nominal voltage in South Africa) and a unity power factor. This calculation entails multiplying the current readings by the nominal voltage, the power factor, and the corresponding time interval. The 2014 dataset includes 1 047 unique profile identifiers corresponding to individual households. Of these, valid readings are available for 786 households. In particular, 5 households devoid of any readings have been excluded from the dataset. Consequently, the analysis is confined to the remaining 781 households. It is worth noting that 7.3% of the consumption dataset has missing values [106], therefore, each household's average daily consumption is determined and used to cluster the load profiles.

6.2 DISAGGREGATING APPLIANCE USAGE FROM METERED HOUSEHOLDS

The clustering of load profiles groups households based on their aggregated electricity consumption from the electricity meters. The aggregated electricity data measures the total power consumption for every appliance at any given time. As the power consumption of each individual appliance is not directly measured, it remains unknown. If we can determine the power consumption per appliance in any household, we can formulate and solve the optimal appliance scheduling problem, as demonstrated in the previous study [64]. Therefore, the following research question arises: How can we disaggregate the metered datasets used for clustering into specific appliance usage schedules for optimal scheduling?

© University of Pretoria

To address this research question, this study investigates non-intrusive load monitoring. Non-intrusive load monitoring (NILM) is a method used to dissagregate total power consumption into individual appliance power consumption without the need for smart meters in each appliance [107]. Instead of placing sensors on each appliance, NILM uses a single or a few meters to measure the overall electricity usage and then disaggregates this data to identify the power consumption patterns of individual appliances. In general, there are three approaches to NILM: 1) pattern matching-based load disaggregation, 2) machine learning-based load disaggregation involving regression or classification models such as hidden Markov models, LSTM and CNN, and 3) source separation-based load disaggregation involving techniques such as sparse component analysis and nonnegative matrix factorisation/nonnegative tensor factorisation [108]. Although NILM inherently deals with single-channel source separation, architectures developed on this basis have not matched the performance of machine learning or pattern matching methods. A major drawback of single-channel source separation techniques is their insufficient incorporation of temporal context and limitations in the use of multidimensional data [109].

Pattern matching in the context of NILM involves comparing the electrical signatures or patterns of total energy consumption with known patterns of individual appliances to identify and disaggregate their contributions. Pattern matching can be implemented using techniques such as dynamic time warping, multivariance matching, and other similarity-based algorithms to accurately identify and separate the energy consumption of different appliances in a household or building [110]. The approach involves comparing the unknown observed appliance signatures with those of known reference signatures to accurately estimate appliance power consumption.

Signatures in NILM are unique patterns or characteristics in the power consumption data of appliances that are specific to each appliance. These signatures help algorithms such as transient power waveforms [111] identify and disaggregate the energy consumption of individual appliances without the need for intrusive monitoring or measurements at the individual appliance level. By analysing features such as power consumption profiles, voltage fluctuations, current patterns, and frequency variations, NILM algorithms can differentiate between appliances according to their energy consumption behaviour. The adaptive factorial hidden Markov model (FHMM) [112] and hybrid signature-based FHMM [113], for example, divide the power signal into a series of different events according to their variation and then use pattern matching to identify different events corresponding to each of the appliances. Combined hidden states are the aggregated states of multiple appliances in the power sequence being analysed,

© University of Pretoria

inferred using the FHMM model, while the real hidden states represent the actual states of individual appliances contributing to the total power consumption. FHMM aims to map the combined hidden states back to the real hidden states of each appliance for an accurate load disaggregation and energy consumption analysis in NILM systems.

Although this study does not use the FHMM method for load disaggregation, it uses the principle of pattern matching from appliance signatures. The key features for the appliance consumption signatures are derived from the DELS survey dataset using the DELSKV appliance variables dataset are as follows:

1. Each household's aggregated meter reading in the DELMH dataset can be matched with its respective aggregated home appliances.
2. The type of appliances owned by each household is provided in the DELS survey datasets and is merged with the DELMH using the DELSKV key variables data.
3. The frequency with which appliances are used by each household is provided in the DELS survey datasets and is merged with the DELMH dataset using the DELSKV key variable data.
4. Individual appliances are not measured directly.

This study assumes that if aggregated meter readings and disaggregated appliance ownership are known, disaggregated power consumption and appliance usage can be determined by matching the pattern of the aggregated load profile based on the baseline appliance schedule in the previous study [64] and the constraint of equal energy that ensures that the integral of the disaggregated load profile and the aggregated metered profile is equal over a specified time interval. Unlike the bottom-up methodology used in a previous study to establish a baseline profile of appliance use from individual appliance consumption [64], this study adopts a top-down approach from total appliance consumption. This method is based on historical metered data (DELMH) merged with appliance ownership and usage frequency survey data (DELS) using key appliance variables (DELSKV). The DELS data lists the appliances (electric stove with three and four plates, hotplate, microwave, washing machine, refrigerator, geyser, electric heater, iron, and kettle) considered when developing the baseline schedules based on the meter readings. However, the exact appliance schedule for the individual appliances related to the aggregated meter readings is unknown.

© University of Pretoria

For example, ProfileID 1006665 appears in both the survey and the metering data. In the DELS survey data, ProfileID 1006665 contains information on a single household in Woodhill in the city of Tshwane with a $390m^2$ floor area, that owns a 4-plate stove, a refrigerator, 3 geysers, 3 heaters, an iron, a kettle, a microwave, 2 television sets and 1 washing machine, with the appliance usage frequency for each appliance. The survey data provides important information about the appliances owned by the household and the frequency of their use. Each appliance is assigned a single power rating based on the typical standard power specifications of the appliances. The same ProfileID 1006665 contains the hourly metered data in the DELMH dataset and is merged with the survey data. Using the on/off control method and assumed power ratings:

1. Set the initial switching status of each appliance based on the baseline schedule in the previous study [64].
2. If the appliance is on, subtract its power usage from the total measured power for the 24-hour duration. Repeat this process for each appliance.
3. After assigning the initial switching status values, the total usage of the appliance for 24 hours is adjusted to ensure that the sum of the disaggregated appliance power is equal to the aggregated measured power.
4. For small variations, adjustments are made by scaling down larger, more flexible appliances, such as a geyser and a heater, as well as scaling down the initially assumed power ratings.
5. The on/off control schedule in [64] ensures that appliances with intermittent usage are not active during hours when they are unlikely to be used.
6. Iterative adjustments are made to the appliance usage, ensuring that the sum of disaggregated appliance loads matches the metered readings.

This approach facilitates both clustering and scheduling strategies. Future research will consider deterministic load signatures for non-intrusive load identification to match power consumption patterns.

6.3 LOAD SHEDDING IN HATFIELD

The load shedding in South Africa has the following key features: 1) the national electricity utility announces that the demand is greater than the dispatchable generation and implements load shedding at short notice as the last resort to prevent grid collapse; 2) the frequency and duration of load shedding are ranked by stages, and the higher the stage, the more severe the load shedding; 3) there is no

© University of Pretoria

forecast of when load shedding will be implemented or how long it will be implemented; This makes it difficult to schedule activities for the next week, month, or even year, when there is no forecast of the exact start and end time of load shedding and when there is no forecast for the exact duration of each outage.

The specific region on which we focus in the case study is Hatfield, South Africa. The annual load shedding hours, represented by the blue axes and bar graph, in Figure 6.1 are derived from historical data on Hatfield's load shedding outages from 2014 to 2023 [64]. Interestingly, after load shedding in 2015, Hatfield did not experience load shedding in 2016 and 2017. Outages increase steadily from 2018 to 2021. There is an exponential increase in 2022 that results in more load shedding than in all previous years combined. The same alarming trend continues in 2023 which has an exponential increase in load shedding, surpassing 2022 in experiencing more load shedding than all previous years combined.

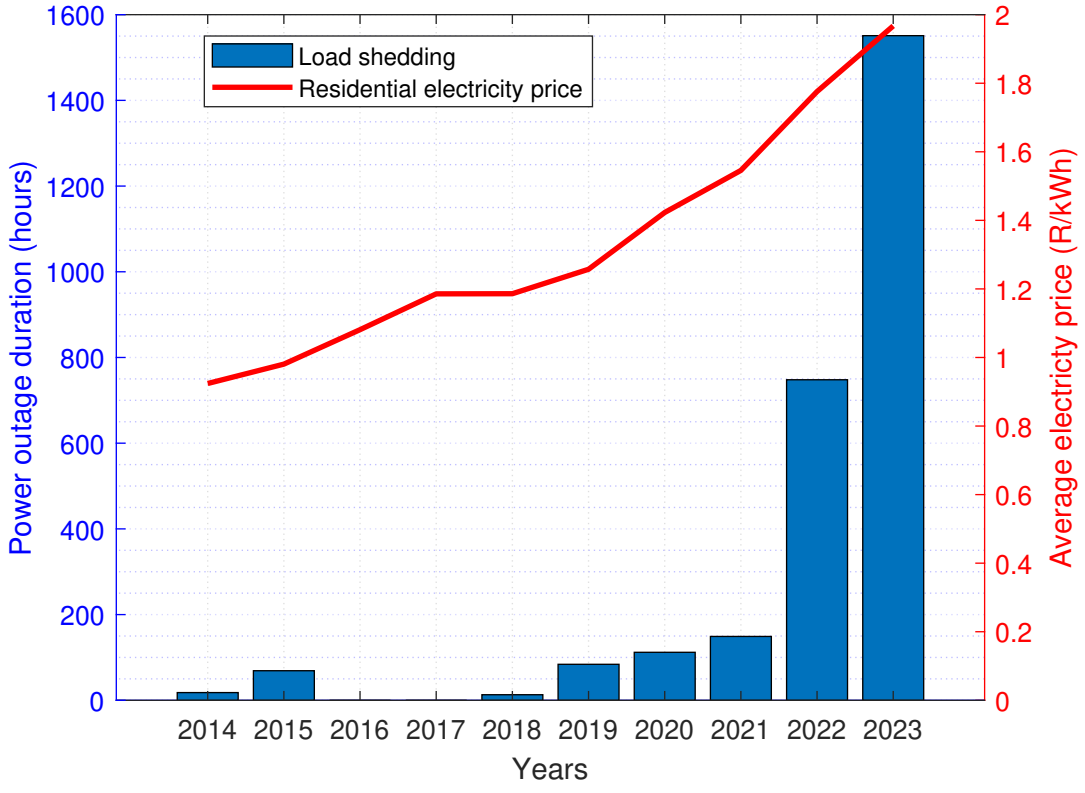

Figure 6.1. The annual load shedding hours and average electricity price in Hatfield, South Africa.

The red axis and the red line graph in Figure 6.1 show the average electricity price for Hatfield residential households from 2014 to 2023, based on the Residential Homelight 20A and 60A incline

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

block tariff. Despite severe load shedding outages during these years, the average residential electricity price has consistently increased over the past decade. Since load shedding began in 2008, Eskom's tariffs have risen by 450% due to declining sales and increasing operating and maintenance costs. Consequently, in 2023, households paid more for less electricity compared to any previous year. This ties into this study's research question by highlighting the dual impact of increasing load shedding hours and subsequent increasing electricity tariffs experienced by households.

To address the research question on the optimal integration of scheduling and sizing for the most convenient strategy to minimise involuntary scheduling from load shedding while effectively integrating solar panels and batteries for various households, this study initiates the results analysis with clustering.

6.4 OPTIMAL NUMBER OF CLUSTERS K, AND OPTIMAL LOAD PROFILE CLUSTERING

In cluster analysis, the Davies-Bouldin index (DBI) is a popular metric used to evaluate the optimal number of clusters [114], where the optimal number is determined based on the clustering algorithm that results in the lowest DBI value. For evaluation, K-means clustering is used as a clustering algorithm in the DBI criterion. Although the DBI index is applicable to various clustering algorithms beyond the K-means, the study by [61] demonstrates that the same optimal number of clusters can be achieved using hierarchical agglomerative clustering, suggesting that using different clustering algorithms would not have a significant impact on the outcome of your DBI value when considering load clustering. The DBI is particularly useful because it combines two key aspects of cluster quality: 1) within-cluster similarity, which refers to the homogeneity within clusters and assesses how closely related the points are within the cluster, and 2) between-cluster separation, which refers to the distinctiveness among clusters and evaluates how well separated a cluster is from other clusters. When determining the most suitable number of clusters, the Davies-Bouldin Index (DBI) plays a crucial role in measuring both the compactness and separation of clusters, with a lower DBI indicating a more desirable cluster structure. In the case of the two clusters shown in Figure 6.2, a DBI of 1.1 suggests a moderate level of cohesion and separation. The optimal configuration is found to have three clusters, where the DBI is minimised to 0.9, indicating improved internal density and clear separation between clusters. However, increasing the number of clusters to four results in a higher DBI of 1.6, indicating a decrease in clustering quality possibly due to over-segmentation, and only slightly improving to 1.4 for five clusters. The six-cluster

[3] www.eskom.co.za/distribution/tariffs-and-charges/

© University of Pretoria

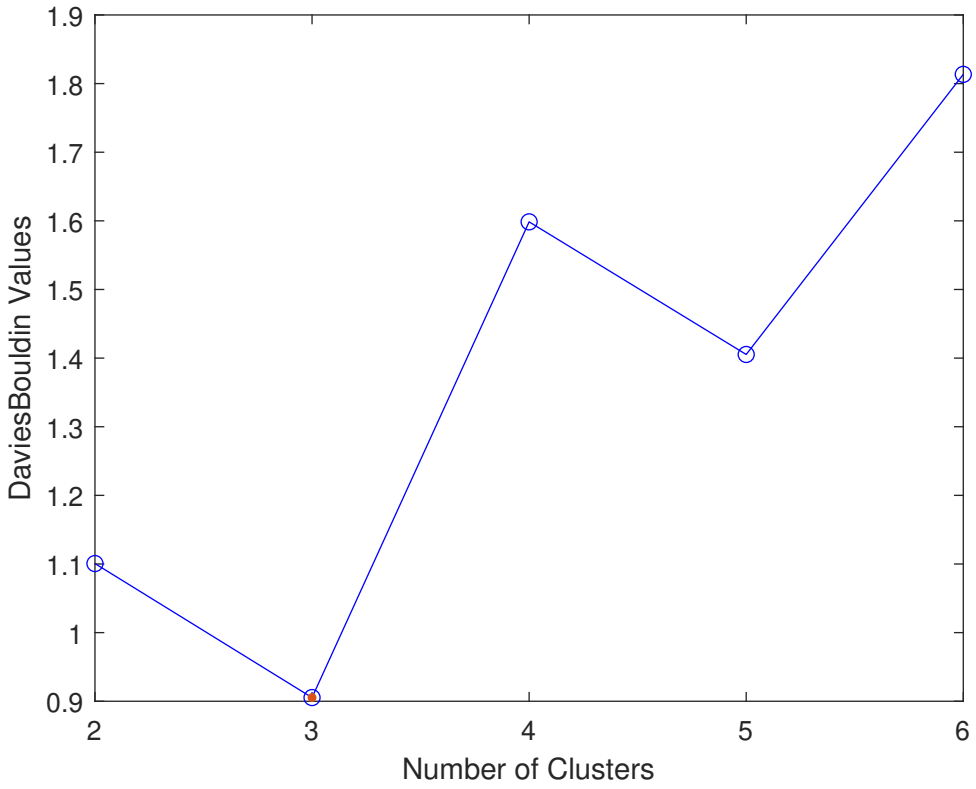

Figure 6.2. Determining the optimal number of clusters K, in K-means clustering, based on the Davies-Bouldin criterion.

archetypes, with the highest DBI of 1.8, represent the least favourable scenario, highlighting the effectiveness of DBI in evaluating cluster quality. The index is computed by averaging the similarity measure across all clusters, where the similarity measure is defined as the ratio of the sum of the scatter within the cluster to the distance between cluster centroids. Analysing the DBI values presented in Figure 6.2 for varying cluster numbers offers valuable information on the quality of the clustering arrangement. For example, a DBI of 1.1 for two clusters suggests a moderate level of cohesion within clusters and separation between clusters, making it a reasonably effective cluster configuration. In contrast, a DBI of 0.9 for three clusters emerges as the lowest among the evaluated configurations, indicating a superior clustering pattern characterised by increased density within clusters and improved distinctiveness among clusters. This implies that a three-cluster configuration is likely the most optimal for the given dataset. On the other hand, a DBI of 1.6 for a four-cluster configuration signifies a decrease in cluster quality, possibly due to the artificial splitting of data points into less cohesive clusters. An incremental enhancement is noticed with five clusters, as shown by a DBI of 1.4. However,

© University of Pretoria

this arrangement still does not reach the optimal structure obtained with three clusters. Lastly, a six-cluster arrangement, denoted by the highest DBI of 1.8, clearly indicates the weakest clustering setup, characterised by reduced coherence and separation between clusters. The DBI curve, which decreases significantly for a low number of clusters before increasing for larger numbers of clusters, is quite similar to the DBI curve in [115]. This similarity persists even though their study used a different dataset: the Low Carbon London project, which includes metered data from 5 567 residential electricity users in London from 2011 to 2014. The silhouette analysis depicted in Figure 6.3 provides

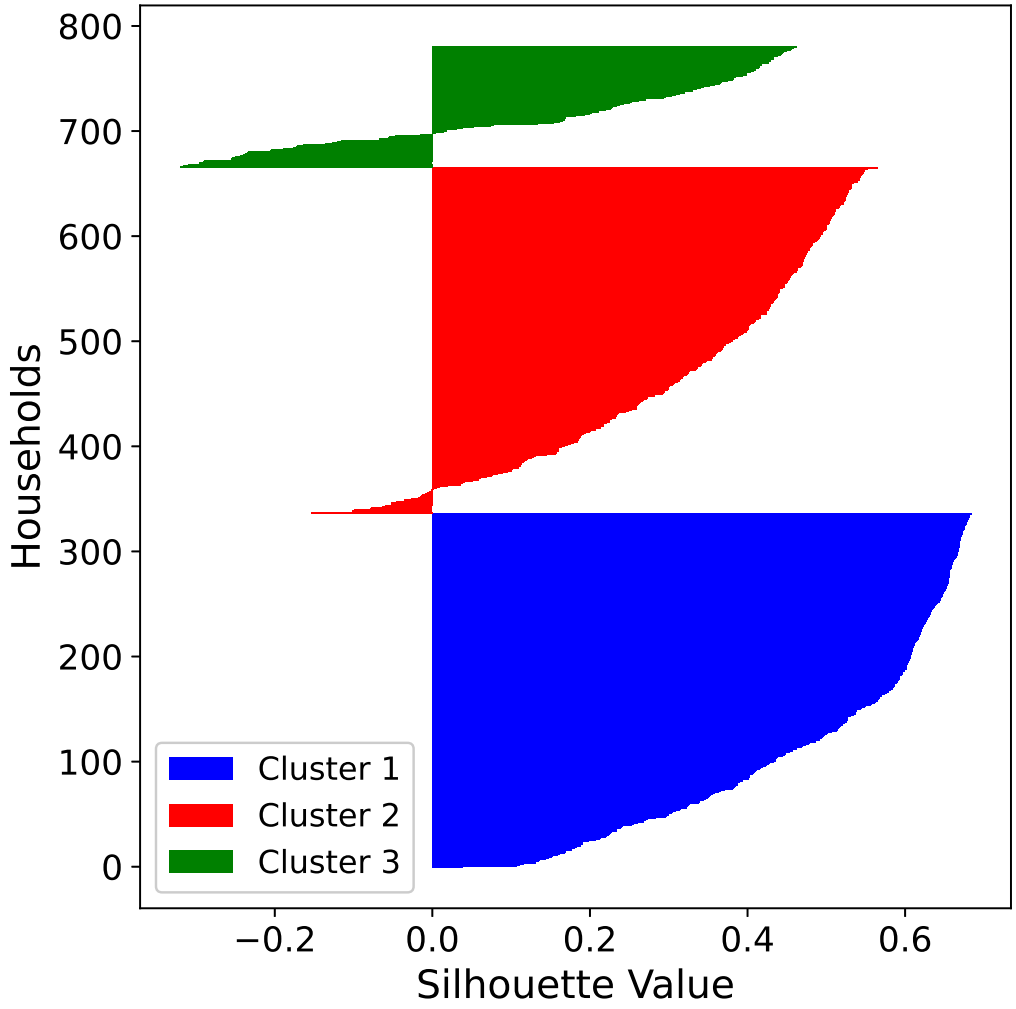

Figure 6.3. The silhouette values assessing the degree of separation among three clusters using the cityblock distance metric for K-means clustering.

a visual representation of how well each data point fits within its cluster, it allows us to determine both the within-cluster similarity and the between-cluster dissimilarity, as determined by a silhouette value ranging from -1 to 1 [116]. The y-axis represents the households after they have been clustered and

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

rearranged such that the households with the lowest silhouette values are at the bottom and households with the highest silhouette values at the top of each cluster. The silhouette value is a measure used to determine how similar an object is to its own cluster compared to other clusters in a dataset. It is primarily used in cluster analysis to assess the quality of the clusters formed. For each data point p, the silhouette value $s(p)$ is calculated by first calculating the average distance to other points within the same cluster $A(p)$ as

$$A(p) = \frac{1}{|C_l| - 1} \sum_{j \in C_l, j \neq p} d(p, j), \tag{6.1}$$

where $|C_l|$ is the number of points in the same cluster as p, and $d(p, j)$ is the distance between points p and j [117]. Secondly, the average distance to points in the nearest (most similar) neighbouring cluster $B(p)$ is determined as

$$B(p) = \min_{C_l' \neq C_l} \frac{1}{|C_l'|} \sum_{j \in C_l'} d(p, j), \tag{6.2}$$

where, C_l' represents any other cluster different from C_l. Therefore, the silhouette value $s(p)$ for each point is formulated as

$$s(p) = \frac{B(p) - A(p)}{\max(A(p), B(p))}, \tag{6.3}$$

where the silhouette value $s(p)$ ranges from -1 to 1:

- $s(p)$ close to 1: The point is well-clustered.
- $s(p)$ close to 0: The point is on or very close to the decision boundary between two neighbouring clusters.
- $s(p)$ close to -1: The point might be assigned to the wrong cluster.

The silhouette value indicates the extent to which a point is correctly assigned to its cluster based on its proximity to other points within the same cluster compared to points in neighbouring clusters. A higher silhouette value suggests a better fit to the cluster, indicating a more precise clustering arrangement. The silhouette plot reveals that the data is divided into three clusters of varying sizes and shapes. Ideally, a silhouette plot would show clusters of equal size and shape, indicating well-separated clusters. Moreover, an ideal silhouette would not contain negative values within any of the clusters. On examination, it is evident that cluster 1 shows a moderate level of alignment with the cluster, as evidenced by a peak silhouette value of 0.45, although with some minor deviations that suggest potential boundary cases. Cluster 2, characterised by a triangular silhouette that peaks at 0.55, demonstrates strong coherence and a clear distinction from neighbouring clusters, indicating effective clustering. However, Cluster 2 does have households that have negative silhouette values. In contrast, cluster 3

© University of Pretoria

exhibits considerable variability in point alignment, with positive and negative peaks suggesting a mixture of well-placed and poorly matched points, indicating areas where clustering accuracy could be improved. This nuanced analysis underscores the intricate dynamics involved in cluster formation and underscores the importance of silhouette analysis in assessing cluster effectiveness. In particular, cluster 2 stands out with the highest silhouette value, indicating its distinctiveness and well-defined nature compared to the other clusters. Cluster 1 shows moderate cohesion and separation, with room for potential improvement, while cluster 3 appears to be the least well defined, with a notable portion of points possibly misaligned or indicating the presence of overlapping boundaries with other clusters. In terms of cluster quality and distance metrics selection, this analysis suggests that while some clusters are clearly defined and cohesive, others may benefit from re-assessment or alternative clustering strategies, such as adjusting parameters or exploring different distance metrics to improve overall clustering performance.

After determining the optimal number of clusters with Davies-Bouldin criteria and evaluating the separation of the clusters using a silhouette plot, the baseline power demand of each of the clusters is determined by analysing Figure 6.4. The blue represents cluster 1, the red cluster 2 and the green cluster 3 in the box-and-whisker plot that provides statistical information about the median in yellow, the quartiles as the edges of the boxes at each hour, and the outliers as coloured dots for reach cluster, respectively.

A diagnostic analysis of the results of the clustering of K means in terms of power consumption in the three clusters, over a 24-hour period in Figure 6.4 reveals distinct profiles of energy use, aligning with the clustering goal of categorising households into low-, medium- and high-energy-consumption clusters. Cluster 1, with a median power usage of 0.18 kW and a mean of 0.194 kW, presents the lowest energy consumption, characterised by the narrowest hourly range (0 kW to 1.69 kW) and fewest outliers. Its interquartile range (IQR) of 0.189 kW suggests a relatively consistent pattern of energy consumption, confirming its classification as the low energy consumption cluster. The mean daily energy consumption of cluster 1 is 4.49 kWh. In contrast, cluster 3 stands out as the cluster of high energy consumption. It presents the highest median (0.911 kW) and mean (1.020 kW) power consumption values, with the widest range (0.163 kW to 4.034 kW) and a substantial number of outliers. The larger IQR of 0.479 kW indicates greater variability, which is consistent with the more diverse or intense energy usage patterns expected in a high consumption load profile. The mean daily energy consumption of cluster 3 is 22.06 kWh. Cluster 2 serves as the intermediate cluster, with

© University of Pretoria

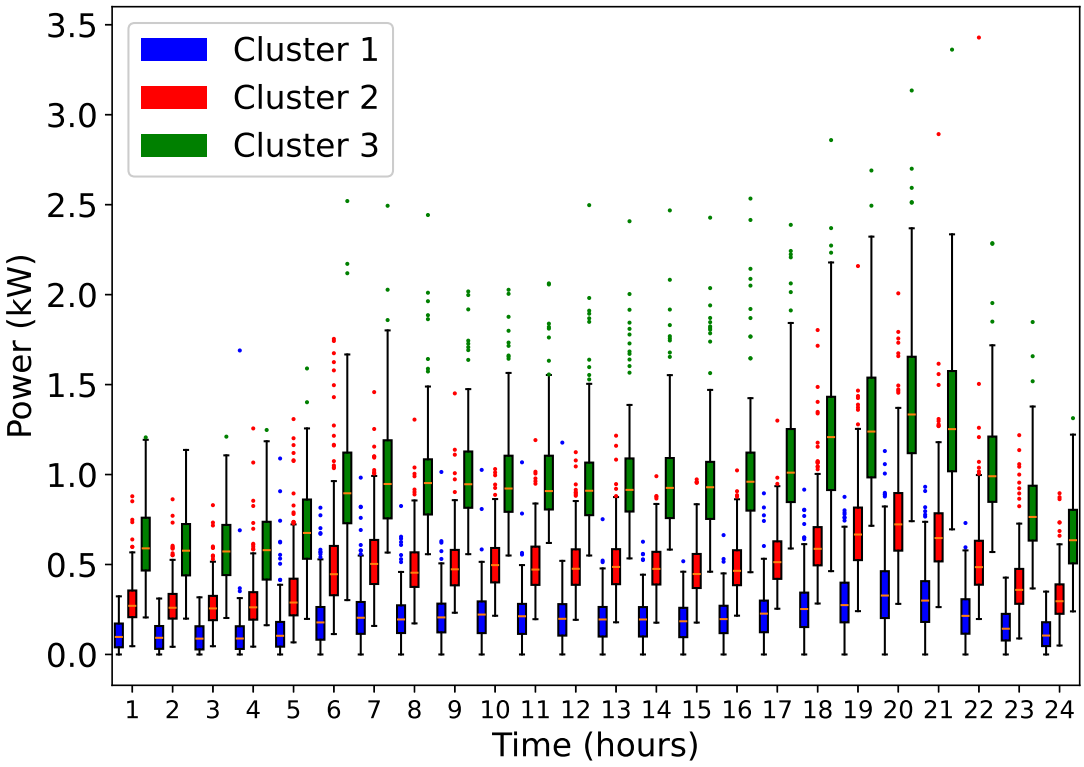

Figure 6.4. The box-and-whisker plot of K-means clustered households shows load profiles classified into low, medium, and high energy consumption clusters 1, 2 and 3, respectively.

median and mean power consumption levels of 0.449 kW and 0.477 kW, respectively. The mean daily energy consumption of cluster 2 is 10.79 kWh. Although cluster 2 has a moderate consumption level, it has the highest number of outliers with values ranging from 0.996 kW to 3.429 kW. This indicates that while the cluster generally represents medium energy usage, there are instances of significantly higher consumption, and much like the negative silhouette values for some households in Figure 6.3 may actually be better suited in cluster 3 and require further investigation, to understand why the unsupervised machine learning algorithm of K-means clustering classified them as cluster 2 instead of cluster 3. After determining the optimal number of clusters with Davies-Bouldin criteria and evaluating the separation of the clusters using a silhouette plot, the baseline power demand of each of the clusters is determined by analysing Figure 6.4. The blue represents cluster 1, the red cluster 2 and the green cluster 3 in the box-and-whisker plot that provides statistical information about the median in yellow, the quartiles as the edges of the boxes at each hour, and the outliers as coloured dots for reach cluster, respectively. A diagnostic analysis of the results of the clustering of K

© University of Pretoria

means in terms of power consumption in the three clusters, over a 24-hour period in Figure 6.4 reveals distinct profiles of energy use, aligning with the clustering goal of categorising households into low-, medium- and high-energy-consumption clusters. Cluster 1, with a median power usage of 0.18 kW and a mean of 0.194 kW, presents the lowest energy consumption, characterised by the narrowest hourly range (0 kW to 1.69 kW) and fewest outliers. Its interquartile range (IQR) of 0.189 kW suggests a relatively consistent pattern of energy consumption, confirming its classification as the low energy consumption cluster. The mean daily energy consumption of cluster 1 is 4.49 kWh. In contrast, cluster 3 stands out as the cluster of high energy consumption. It presents the highest median (0.911 kW) and mean (1.020 kW) power consumption values, with the widest range (0.163 kW to 4.034 kW) and a substantial number of outliers. The larger IQR of 0.479 kW indicates greater variability, which is consistent with the more diverse or intense energy usage patterns expected in a high consumption load profile. The mean daily energy consumption of cluster 3 is 22.06 kWh. Cluster 2 serves as the intermediate cluster, with median and mean power consumption levels of 0.449 kW and 0.477 kW, respectively. The mean daily energy consumption of cluster 2 is 10.79 kWh. Although cluster 2 has a moderate consumption level, it has the highest number of outliers with values ranging from 0.996 kW to 3.429 kW. This indicates that while the cluster generally represents medium energy usage, there are instances of significantly higher consumption, and much like the negative silhouette values for some households in Figure 6.3 may actually be better suited in cluster 3 and require further investigation, to understand why the unsupervised machine learning algorithm of K-means clustering classified them as cluster 2 instead of cluster 3.

In addition, the number of households in each of the 3 clusters is provided in Table 6.1, where 43% of the households are clustered in the low-energy-consuming household archetype, 42% in the medium energy and 15% in the high-energy archetype.

Table 6.1. The number of households in each of the K-means clusters.

Clusters	Number of households	% of total households
1	338	43.2778
2	328	41.9974
3	115	14.7247

Figure 6.5 is a histogram of average daily energy consumption in the 3 clusters, divided into 20 bins

© University of Pretoria

for each cluster; the bins that overlap between cluster 1 in blue and cluster 2 in red are highlighted by the subplot in the upper right corner of Figure 6.5.

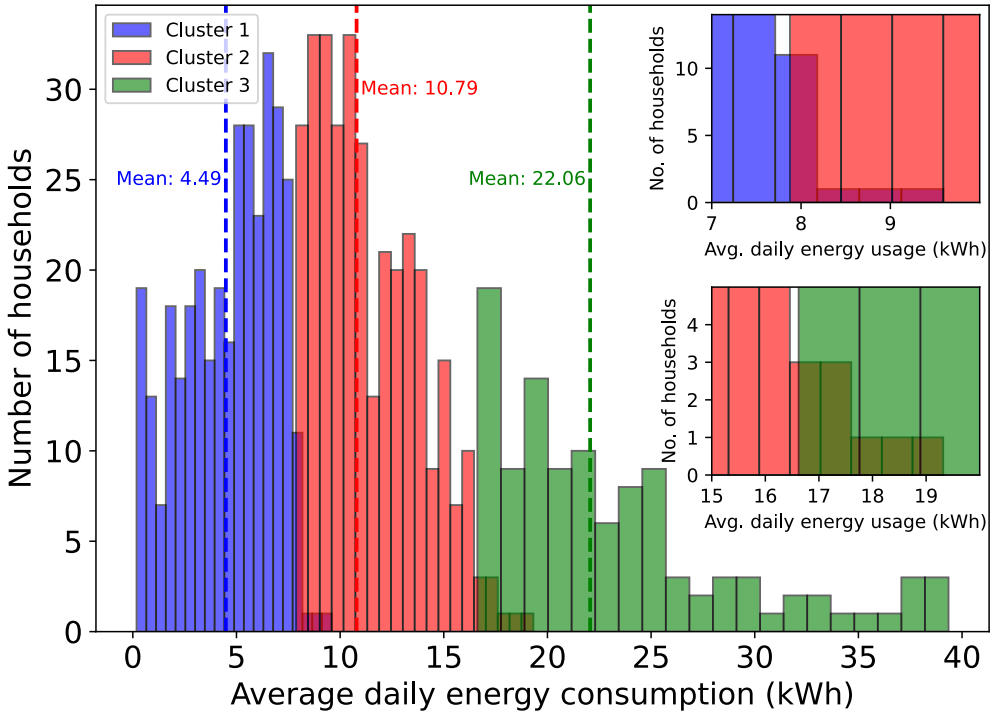

Figure 6.5. A histogram of average daily energy consumption in K-means clusters, divided into 20 bins, with some bins overlapping with other clusters. Blue is for low-energy households, red for medium-energy-consuming households, and green for high-energy households.

The mean for the average daily energy consumption for cluster 1's 338 households is shown with the blue dashed line as 4.49 kWh. The 3 overlapping red bins coincide with the negative silhouette values in Figure 6.3. The red bins for cluster 2 overlap the blue bins for cluster 1 when the average daily energy consumption of low-energy households is greater than or equal to 7.75 kWh. There are 28 households in the first red bin (8.54% of the households in cluster 2) with average daily energy consumption values ranging from 7.75 kWh to 8.4 kWh that can be clustered as low energy consumption households. There are 66 households in the second and third red bins (20.12% of the households in cluster 2) with average daily energy consumption values ranging from 8.4 kWh to 9.6 kWh that can be clustered as households with low energy consumption.

The mean of the average daily energy consumption for the 328 households in cluster 2 is shown with

© University of Pretoria

the dashed red line as 10.79 kWh. The 3 overlapping green bins coincide with the negative silhouette values in Figure 6.3. The green bins for cluster 3 overlap the red bins for cluster 2 when the average daily energy consumption of households with medium energy consumption is greater than or equal to 16.6 kWh. There are 19 households in the first green bin (16.52% of the households in cluster 3) with average daily energy consumption values ranging from 16.6 kWh to 17.75 kWh that can be clustered as households with medium energy consumption. There are 9 and 13 households in the second and third red bins, respectively, (7.83% and 11.30% of the households in cluster 3) with average daily energy consumption values ranging from 17.6 kWh to 18.9 kWh and from 18.9 kWh to 19.3 kWh, respectively, that can be clustered as households with medium energy consumption. The mean of the average daily energy consumption for the 115 households in cluster 3 is shown with the dashed green line as 22.06 kWh.

6.5 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR LOW ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING

If the optimal sizing of a solar panel array is based on its cross-sectional area under predetermined solar irradiance and ambient temperature without considering scheduling, there is a significant risk of oversizing the solar array. Currently, solar arrays around the world are oversized by a factor between 120% and 140% and one of the main reasons why designers oversize solar panel arrays is to provide a contingency margin for peak power generation that ensures peak demand is met [118]. This has a significant financial impact, which often discourages households from adopting solar home strategies. For example, a household may find that without optimal scheduling, the required investment in solar panels significantly exceeds their budget. This happens because the array is sized to meet peak demand scenarios that may only occur a few times a year, leading to higher costs for this capacity, which is not needed most of the time.

Furthermore, during periods of load shedding, households can prioritise the operation of essential appliances through scheduling, rather than operating all appliances simultaneously. Consequently, scheduling facilitates more cost-effective sizing strategies, as it mitigates the need to continuously satisfy the full load profile. Specifically, scheduling the average load profile for cluster 1, which has a daily energy consumption of 4.49 kWh, as part of the sizing strategy for the cross-sectional area of the solar panel array A_c and the battery capacity C_b, considered as decision variables, produces an optimal configuration of a 2.4 kWh battery and a 0.39 kWp solar panel array when the objective function is to minimise both the loss of power supply and inconvenience.

© University of Pretoria

Without incorporating scheduling to minimise inconvenience and power supply loss in the sizing strategy, the solar panel array size would be $1.4 \times P_{cluster1}^{peak} = 1.4 \times 0.3532$ kWp = 0.4945 kWp, and the battery capacity would be $1.4 \times SOE_{cluster1}^{mean} = 1.4 \times 4.49$ kWh = 6.286 kWh, following the general rule of thumb for sizing solar panel arrays [118] and the DC to AC ratio for inverters [119]. Therefore, not considering scheduling in the sizing strategy results in the solar panel array being 51.05% larger than necessary and the battery capacity being 28.57% oversized.

The monthly effects of South Africa's load shedding stages on low-energy-consuming households without solar panels or batteries are shown in Table 6.2. The monthly demand is determined from the cluster 1 centroid in Figure 6.6.

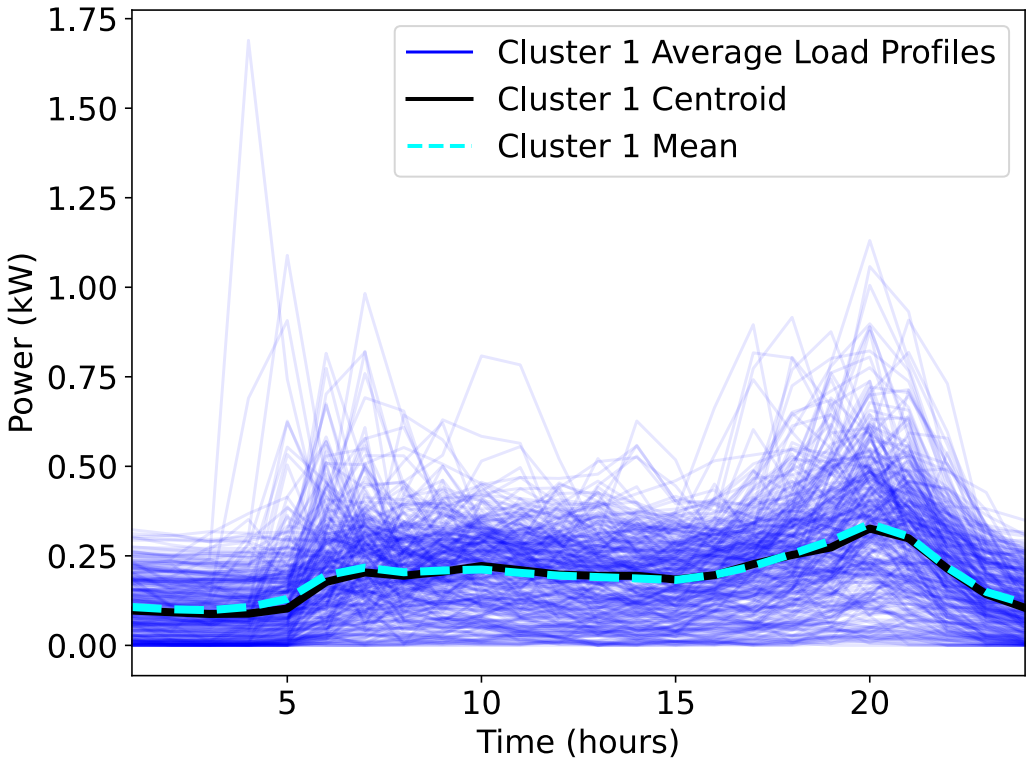

Figure 6.6. The average daily load profiles for cluster 1 households using K-means clustering.

In this instance, the cluster 1 centroid and the cluster 1 mean have the same average daily load profile. The 338 cluster 1 average daily load profiles are represented in blue in Figure 6.6. There are two peaks in the daily household energy consumption profiles for cluster 1's low-energy consumption households, the first in the morning and the second in the evening. When there is no load shedding (stage 0), the

© University of Pretoria

monthly demand for the average low energy consuming household is 139.11 kWh with no loss of power supply and no inconvenience. The loss of power supply and inconvenience for each of the eight stages of load shedding from stage 1 to stage 8 are provided in Table 6.2, with a 6.08% loss of power supply and 4.21 hours of inconvenience at stage 1, to 48.43% loss of power supply and 39.21 hours of inconvenience at stage 8.

Table 6.2. Monthly effects of South Africa's load shedding stages on low energy consuming households without solar panels or batteries.

Stage	Demand (kWh)	Grid supply (kWh)	Loss of power supply (kWh)	(%)	Inconvenience (hours)
0	139.11	139.11	0.00	0.00	0.00
1	139.11	130.65	8.46	6.08	4.21
2	139.11	122.07	17.05	12.25	10.30
3	139.11	113.37	25.74	18.51	15.88
4	139.11	104.40	34.72	24.96	22.73
5	139.11	96.22	42.89	30.83	26.79
6	139.11	88.18	50.93	36.61	30.85
7	139.11	80.16	58.95	42.38	34.65
8	139.11	71.74	67.37	48.43	39.21

Based on the average load profile, which is the mean of cluster 1 in Figure 6.6 and the centroid in Figure 6.7 (a), the inconvenience and loss of power supply caused by load shedding subject to the constraints presented in [64], are minimised in a multi-objective mixed integer nonlinear optimisation problem with weighing factors of $w_1 = 0.5$ and $w_2 = 0.5$, where the weighted sum method is used and the sum of the two weighting factors of the objective function must be equal to one as a necessary condition to have an optimal solution [50]. The aggregated centroid shown in black in Figure 6.7 (a) is disaggregated and shown in blue in Figure 6.7 (a) based on the appliance schedule of the baseline in [64] and the constraint that the integral of the disaggregated and the centroid power curves for 24 hours is equal.

As a result, scheduling the average load profile of cluster 1 with an optimally sized 2.4 kWh battery and 0.39 kWp panel array minimises the power loss caused by load shedding to zero and reduces the involuntary inconvenience by 86.73% and reduces the inconvenience of rescheduling by 93.36%.

© University of Pretoria

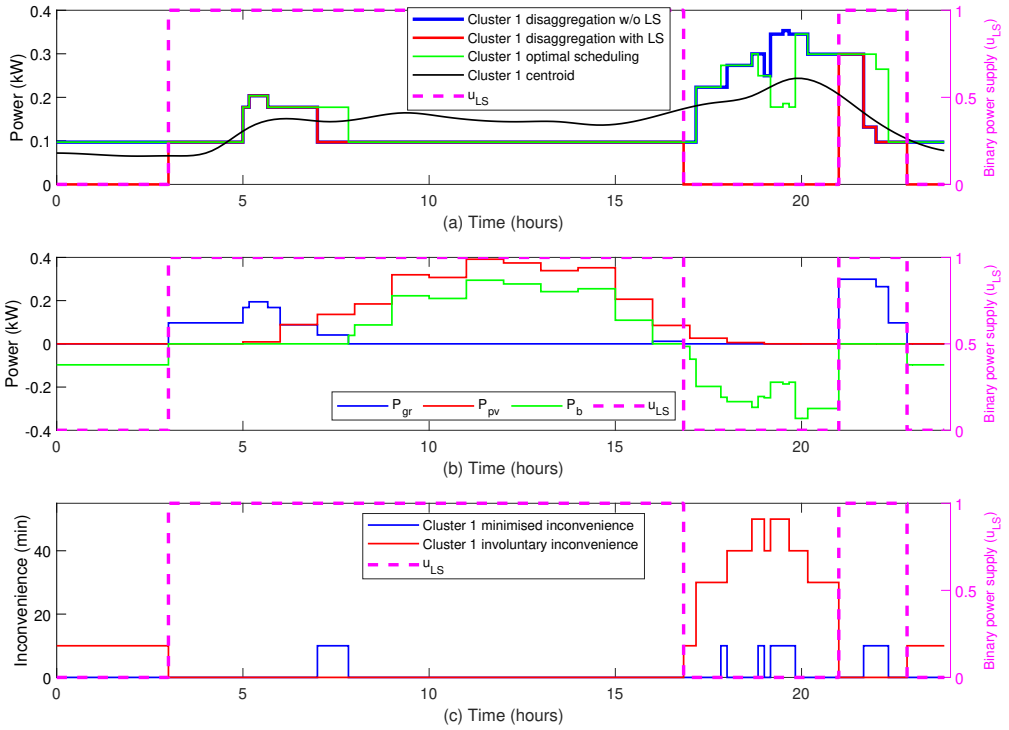

Figure 6.7. (a) The optimal scheduling of cluster 1's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling.

The optimal schedule for appliances that use batteries and solar panels is shown in Figure 6.7 (a), with the optimal power flow management of optimally sized solar and battery energy sources in Figure 6.7 (b).

The inconvenience, both the minimised inconvenience and the involuntary inconvenience, is presented in Figure 6.7 (c) in blue and red, respectively, and clearly shows two periods of inconvenience: 1) the first is during a load shedding power outage when $u_{LS} = 0$ and residents cannot use their non-critical appliances, despite having an optimally sized solar home system; 2) the second is when there is no load shedding power outage when $u_{LS} = 1$ and residents are forced to reschedule their appliances involuntarily to ensure that they can still meet their daily demand.

© University of Pretoria

6.6 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR MEDIUM ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING

If the optimal sizing of a solar panel array for medium-energy-consuming households is based on its cross-sectional area under predetermined solar irradiance and ambient temperature without considering scheduling, there is a significant risk of oversizing the solar array. Furthermore, during periods of load shedding, households can prioritise the operation of essential appliances through scheduling, rather than operating all appliances simultaneously. Consequently, scheduling facilitates more cost-effective sizing strategies, as it mitigates the need to continuously satisfy the full load profile. Specifically, scheduling the average load profile for cluster 2, which has a daily energy consumption of 10.79 kWh when the cluster 2 mean curve in Figure 6.8 is integrated for 24 hours, as part of the sizing strategy for the cross-sectional area of the solar panel array A_c and the battery capacity C_b, considered decision variables, produces an optimal configuration of a 4.8 kWh battery and a 0.78 kWp solar panel array when the objective function is to minimise both loss of power supply and inconvenience.

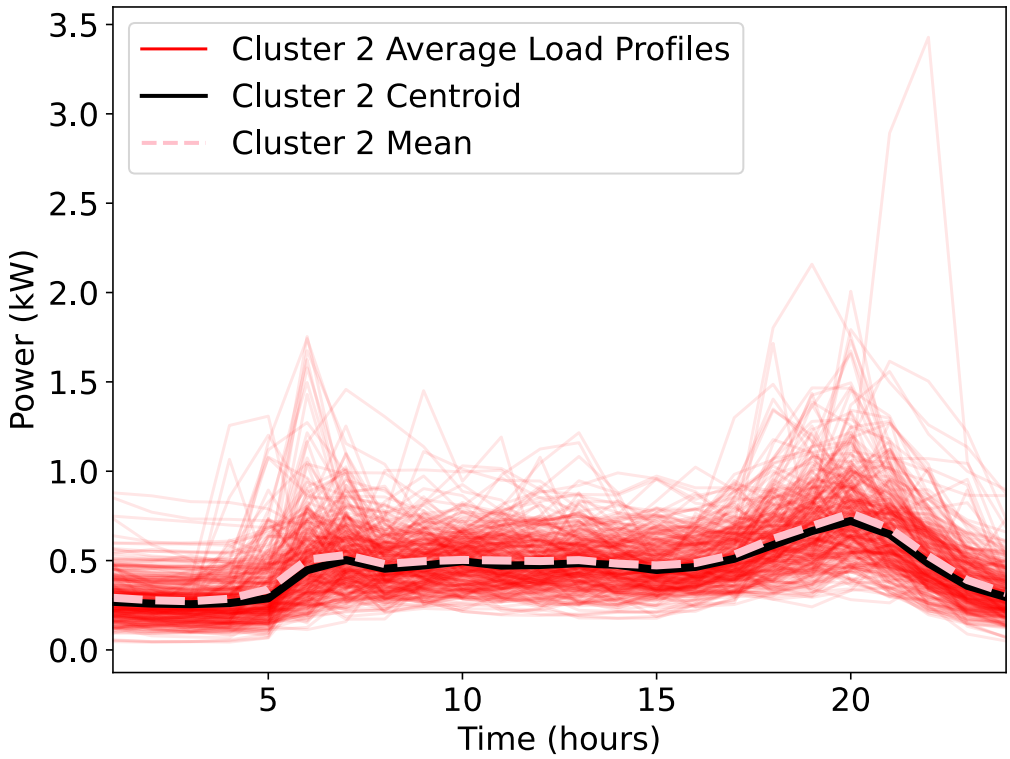

Figure 6.8. The average daily load profiles for cluster 2 households using K-means clustering.

© University of Pretoria

Without incorporating scheduling to minimise inconvenience and power supply loss in the sizing strategy, the solar panel array size would be $1.4 \times P^{peak}_{cluster2} = 1.4 \times 0.8537$ kWp = 1.1952 kWp, and the battery capacity would be $1.4 \times SOE^{mean}_{cluster2} = 1.4 \times 10.79$ kWh = 15.106 kWh, following the general rule of thumb for sizing solar panel arrays [118] and the DC to AC ratio for inverters [119]. Therefore, not considering scheduling in the sizing strategy results in the solar panel array being 34.74% larger than necessary and the battery capacity being 68.22% oversized.

The monthly effects of South Africa's load shedding stages on medium-energy-consuming households without solar panels or batteries are shown in Table 6.3. The monthly demand is determined from the cluster 2 centroid in Figure 6.8. In this instance, the cluster 2 centroid and the cluster 2 mean have the same average daily load profile. The 328 cluster 2 average daily load profiles are represented in red in Figure 6.8.

Table 6.3. Monthly effects of South Africa's load shedding stages on medium energy consuming households without solar panels or batteries.

Stage	Demand	Grid supply	Loss of power supply		Inconvenience
	(kWh)	(kWh)	(kWh)	(%)	(hours)
0	334.61	334.61	0.00	0.00	0.00
1	334.61	314.15	20.46	6.11	10.13
2	334.61	293.52	41.08	12.28	24.77
3	334.61	272.67	61.94	18.51	38.20
4	334.61	251.08	83.52	24.96	54.68
5	334.61	231.72	102.88	30.75	64.43
6	334.61	212.62	121.98	36.46	74.20
7	334.61	193.56	141.04	42.15	83.35
8	334.61	173.55	161.06	48.13	94.33

There are two peaks in the daily household energy consumption profiles for cluster 2's medium-energy consumption households, the first in the morning and the second in the evening. When there is no load shedding (stage 0), the monthly demand for the average medium energy consuming household is 334.61 kWh with no loss of power supply and no inconvenience. The loss of power supply and inconvenience for each of the eight stages of load shedding from stage 1 to stage 8 are provided in Table 6.3, with a 6.11% loss of power supply and 10.13 hours of inconvenience at stage 1, up to 48.13%

© University of Pretoria

loss of power supply and 94.33 hours of inconvenience at stage 8. Based on the average load profile, which is the mean of cluster 2 in Figure 6.8 and the centroid in Figure 6.9 (a), the inconvenience and loss of power supply caused by load shedding subject to the constraints presented in [64], are minimised in a multi-objective mixed integer nonlinear optimisation problem with weighing factors of $w_1 = 0.5$ and $w_2 = 0.5$.

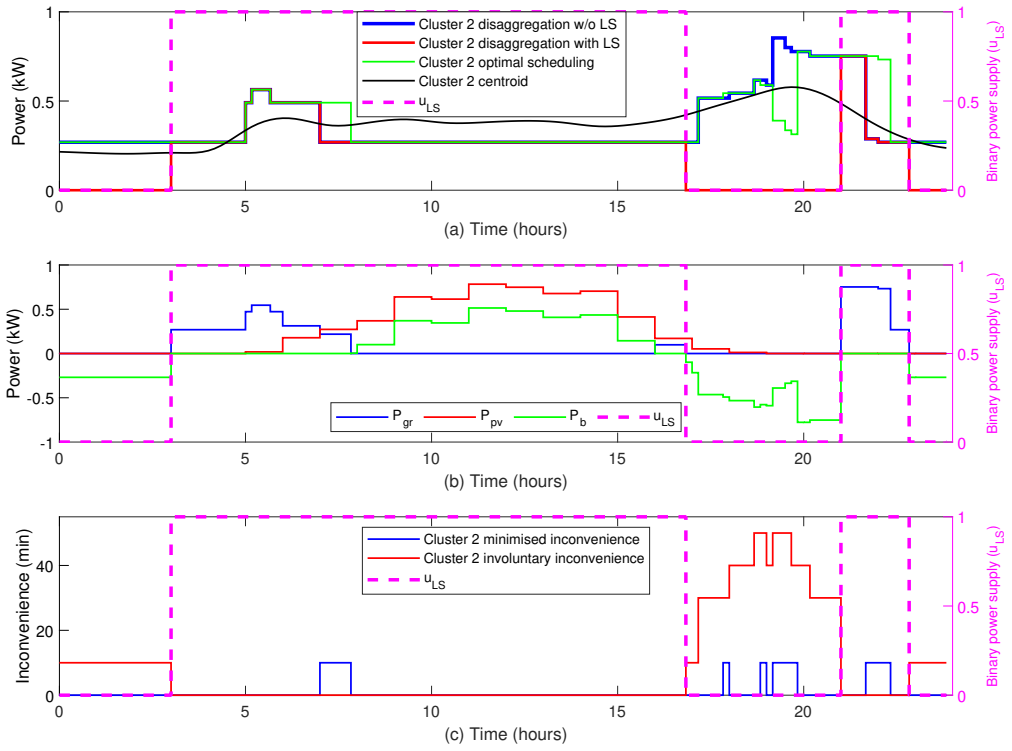

Figure 6.9. (a) The optimal scheduling of cluster 2's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling.

The aggregated centroid shown in black in Figure 6.9 (a) is disaggregated and shown in blue in Figure 6.9 (a) based on the appliance schedule of the baseline in [64] and the constraint that the integral of the disaggregated and the centroid power curves for 24 hours is equal. As a result, scheduling the average load profile of cluster 2 with an optimally sized 4.8 kWh battery and 0.78 kWp panel array minimises the power loss caused by load shedding to zero and reduces the involuntary inconvenience by 86.7% and the inconvenience of involuntary rescheduling by 93.4%. The optimal schedule for appliances that use batteries and solar panels is shown in Figure 6.9 (a), with the optimal power flow management of optimally sized solar and battery energy sources in Figure 6.9 (b). In this

© University of Pretoria

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

study, the disparity in household energy consumption explored in [120] between cities is investigated between low-energy and medium-energy consumption households. The disparity in inconvenience in Table 6.2 and Table 6.3 shows that for stage 1 load shedding, low-energy consumption households are inconvenienced for 4.21 hours, while medium energy consumption households are inconvenienced for 10.13 hours. Therefore, households with medium energy consumption are inconvenienced 2.4062 times more than households with low energy consumption.

6.7 OPTIMAL SCHEDULING WITH SOLAR AND BATTERY SIZING FOR HIGH ENERGY CONSUMING HOUSEHOLDS UNDER LOAD SHEDDING

If the optimal sizing of a solar panel array for high-energy-consuming households is based on its cross-sectional area under predetermined solar irradiance and ambient temperature without considering scheduling, there is a significant risk of oversizing the solar array. Furthermore, during periods of load shedding, households can prioritise the operation of essential appliances through scheduling, rather than operating all appliances simultaneously. Consequently, scheduling facilitates more cost-effective sizing strategies, as it mitigates the need to continuously satisfy the full load profile. Specifically, scheduling the average load profile for cluster 3, which has a daily energy consumption of 22.06 kWh when the cluster 3 mean curve in Figure 6.10 is integrated over 24 hours, as part of the sizing strategy for the cross-sectional area of the solar panel array A_c and the battery capacity C_b, considered decision variables, produces an optimal configuration of a 7.2 kWh battery and a 1.17 kWp solar panel array when the objective function is to minimise both the loss of power supply and inconvenience.

Without incorporating scheduling to minimise inconvenience and power supply loss in the sizing strategy, the solar panel array size would be $1.4 \times P^{peak}_{cluster3} = 1.4 \times 1.71$ kWp = 2.39 kWp, and the battery capacity would be $1.4 \times SOE^{mean}_{cluster3} = 1.4 \times 22.06$ kWh = 30.88 kWh, following the general rule of thumb for sizing solar panel arrays [118] and the DC to AC ratio for inverters [119]. Therefore, not considering scheduling in the sizing strategy results in the solar panel array being 51.05% larger than necessary and the battery capacity being 28.56% oversized.

The monthly effects of South Africa's load shedding stages on low-energy-consuming households without solar panels or batteries are shown in Table 6.4. The monthly demand is determined from the cluster 3 centroid in Figure 6.10. In this instance, the cluster 3 centroid and the cluster 3 mean have the same average daily load profile. The 115 cluster 3 average daily load profiles are represented in green in Figure 6.10. There are two peaks in the daily household energy consumption profiles for

© University of Pretoria

cluster 3's high-energy consumption households, the first in the morning and the second in the evening. When there is no load shedding (stage 0), the monthly demand for the average high energy consuming household is 683.84 kWh with no loss of power supply and no inconvenience. The loss of power supply and inconvenience for each of the eight stages of load shedding from stage 1 to stage 8 are provided in Table 6.4, with a 6.12% loss of power supply and 20.70 hours of inconvenience at stage 1, to 47.93% loss of power supply and 192.77 hours of inconvenience in stage 8.

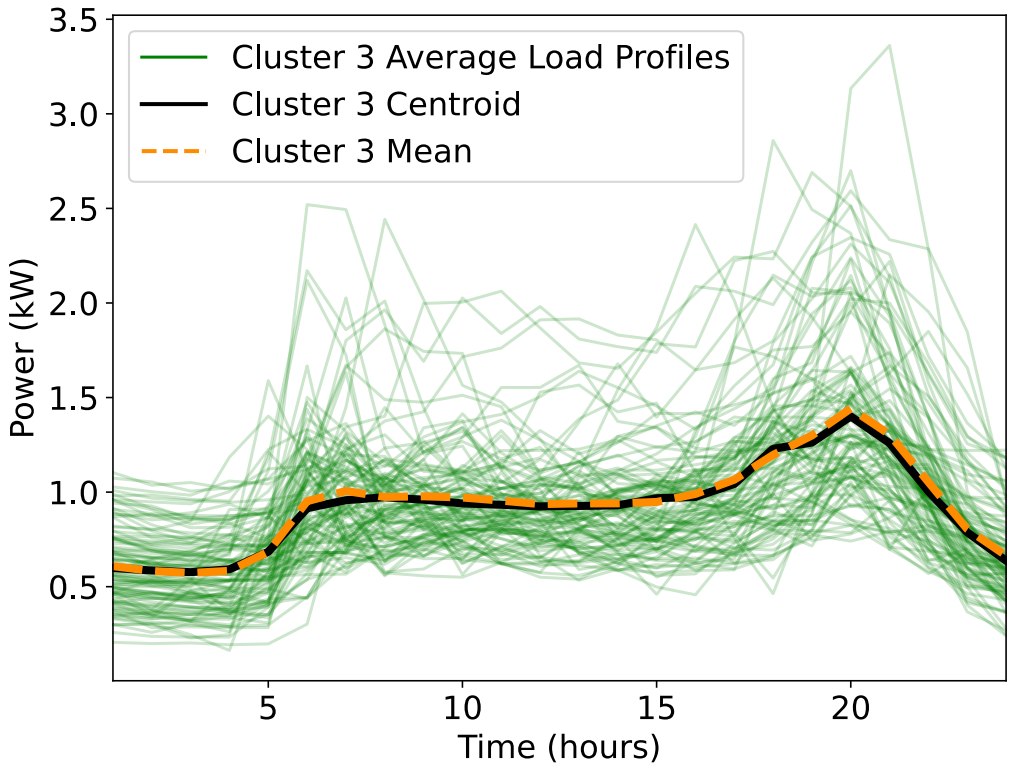

Figure 6.10. The average daily load profiles for cluster 3 households using K-means clustering.

The daily energy consumption of the mean profile of cluster 3 is obtained by integrating the blue power load curve over 24 hours in Figure 6.11 (a), resulting in 22.06 kWh, used as a baseline for high household energy consumption without load shedding. Integrating the red power load curve representing the resultant load profile during load shedding, which matches the blue curve when $u_{LS}(k) = 1$ and is zero when $u_{LS}(k) = 0$, gives 12.67 kWh. Thus, high-energy households lose 9.39 kWh (42.6%) during load shedding. This loss is minimised in the multi-objective function (5.11). The optimal load profile, shown in green in Figure 6.11 (a), restores the daily consumption to 22.06 kWh, minimising the daily loss of power supply from 9.39 kWh to 0 kWh.

© University of Pretoria

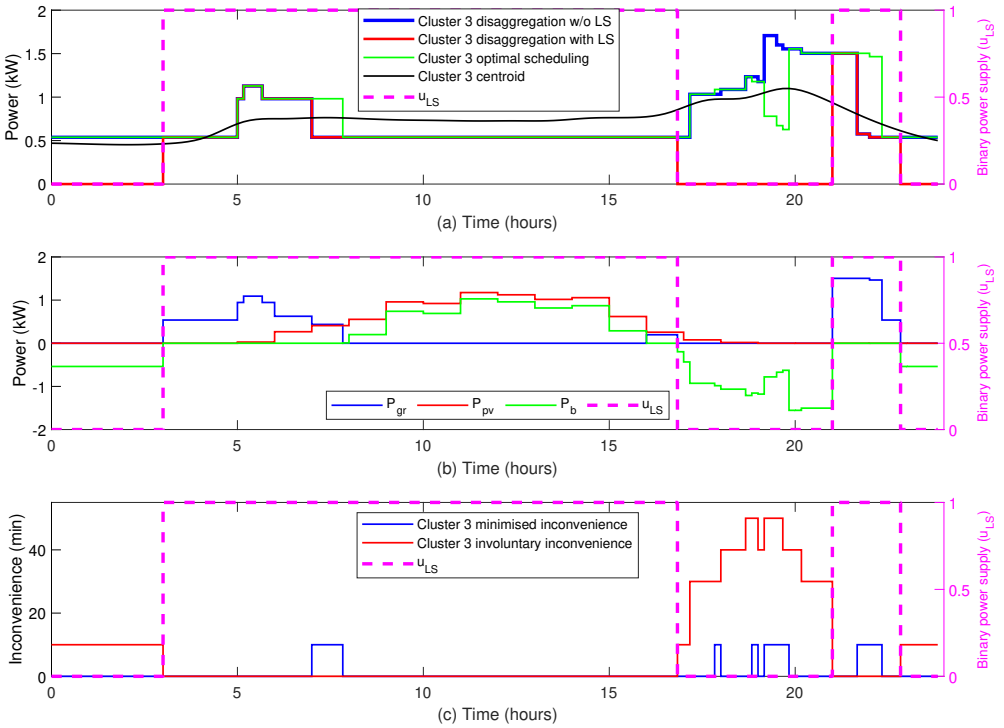

Figure 6.11. (a) The optimal scheduling of cluster 3's average load profile considering load shedding ($u_{LS}(k) = 0$), (b) with the optimal power flow management of optimally sized solar and battery energy sources, (c) and the inconvenience of rescheduling.

Based on the average load profile, which is the mean of cluster 3 in Figure 6.10 and the centroid in Figure 6.11 (a), the inconvenience and loss of power supply caused by load shedding subject to the constraints presented in [64], are minimised in a multi-objective mixed integer nonlinear optimisation problem with weighing factors of $w_1 = 0.5$ and $w_2 = 0.5$. The aggregated centroid shown in black in Figure 6.11 (a) is disaggregated and shown in blue in Figure 6.11 (a) based on the appliance schedule of the baseline in [64] and the constraint that the integral of the disaggregated and the centroid power curves for 24 hours (energy) is equal. As a result, scheduling the average load profile of cluster 3 with an optimally sized 7.2 kWh battery and 1.17 kWp panel array minimises the power loss caused by load shedding to zero and reduces the involuntary inconvenience by 86.4%.

© University of Pretoria

Table 6.4. Monthly effects of South Africa's load shedding stages on high energy consuming households without solar panels or batteries.

Stage	Demand (kWh)	Grid supply (kWh)	Loss of power supply (kWh)	(%)	Inconvenience (hours)
0	683.84	683.84	0.00	0.00	0.00
1	683.84	641.98	41.86	6.12	20.70
2	683.84	599.72	84.12	12.30	50.62
3	683.84	557.20	126.64	18.52	78.07
4	683.84	513.08	170.76	24.97	111.75
5	683.84	473.93	209.91	30.70	131.68
6	683.84	435.29	248.55	36.35	151.65
7	683.84	396.64	287.20	42.00	170.33
8	683.84	356.056	327.79	47.93	192.77

6.8 MINIMISING INVOLUNTARY INCONVENIENCE FOR CLUSTERED RESIDENTIAL HOUSEHOLDS UNDER LOAD SHEDDING

The optimised scheduling for all appliances in the clusters is shown in the optimal load profile in green, and the mean of each of the clusters is shown in the demand without load shedding in blue (Figures 6.7(a), 6.9(a) and 6.11(a)). These optimally scheduled appliances are based on the average load profiles within the low, medium and high energy-consuming clusters (Figures 6.6, 6.8, and 6.10).When the mean and optimal profiles are the same, the inconvenience is zero, as shown by the zero inconvenience values (Figures 6.7(c), 6.9(c) and 6.11(c)). The interpretation in that case is that the household is scheduled to use its appliances no differently than how the appliances are used ordinarily or on average when there is no load shedding. When the optimal profile does not match the mean profile, the household is inconvenienced.

1. Involuntary inconvenience in the morning (00:00 - 02:50): During the load shedding outage in the morning from 00:00 to 02:50 when $u_{LS}(k) = 0$, households with low, medium and high energy consumption profiles, which have not invested in alternative power strategies, experience a complete loss of grid power and lack any backup power provisions during this period. Consequently, this study identifies two quantifiable impacts experienced by households. a) Without

© University of Pretoria

any intervention, households endure involuntary inconveniences, forfeiting the opportunity to use their appliances during the outage. As illustrated by the red curve in Figure 6.7 (c), Figure 6.9 (c), and Figure 6.11 (c) households face 170 minutes of involuntary inconvenience during the outage. b) Should households opt to operate their appliances regardless, they are compelled to reschedule their usage to less convenient times outside the scheduled outages. The appliance shown in Figure 6.7 (c), Figure 6.9 (c), and Figure 6.11 (c) which could not be used during the outage, can be rescheduled to a period outside the outage, though this involuntary rescheduling imposes a 170-minute inconvenience of its own. This study aims to minimise both involuntary inconvenience and involuntary rescheduling, with the outcomes depicted by the blue curve in Figure 6.7 (c), Figure 6.9 (c) and Figure 6.11 (c). The loss of power supply for cluster 1 is 0.28 kWh, for cluster 2 is 0.76 kWh and for cluster 3 is 1.52 kWh during the morning load shedding outage.

2. Minimised inconvenience before the load shedding outage (07:00 - 07:50): The initial interval of minimised inconvenience commences at 07:00 and concludes at 07:50. This interval falls outside the load shedding outage, as indicated by the secondary y-axis on the right side of the plot, where $u_{LS}(k) = 1$. The average household with high energy consumption experiences a 50-minute inconvenience period in the morning from 07:00 to 07:50. This arises because appliances that would typically be used during the load shedding outage are optimally rescheduled to a time outside of this outage, aligning as closely as possible with their standard consumption patterns. In this case, the optimal strategy to minimise inconvenience involves rescheduling one appliance to operate 50 minutes earlier in the morning. In scenarios where multiple appliances require rescheduling within the same time frame, the inconvenience is amplified proportionately to the number of appliances requiring operation to mitigate the loss of power and associated inconvenience. For example, should two appliances be rescheduled for the given period, the corresponding value in Figure 6.7 (c), Figure 6.9 (c) and Figure 6.11 (c) would increase to 20, thereby extending the inconvenience to 100 minutes, or an hour and 40 minutes, compared to the initial 50 minutes. The duration of inconvenience for each appliance is determined by the time it is operated outside its usual usage schedule. The optimally rescheduled appliance for cluster 1 requires 0.1772 kW, with 0.1362 kW from photovoltaic solar panels $P_{pv}(k)$ and 0.0410 kW from the electrical grid $P_{gr}(k)$. The battery $P_b(k)$ in cluster 1 neither charges nor discharges when there is grid power (when $P_{gr}(k) > 0$). The optimally rescheduled appliance for cluster 2 requires 0.4907 kW, with 0.2724 kW from photovoltaic solar panels $P_{pv}(k)$ and 0.2183 kW from the electrical grid $P_{gr}(k)$. The battery $P_b(k)$ in cluster 2 neither charges nor discharges

© University of Pretoria

when there is grid power (when $P_{gr}(k) > 0$). The optimally rescheduled appliance for cluster 3 requires 0.9815 kW, with 0.4086 kW from photovoltaic solar panels $P_{pv}(k)$ and 0.4367 kW from the electrical grid $P_{gr}(k)$. The battery $P_b(k)$ in cluster 3 neither charges nor discharges when there is grid power (when $P_{gr}(k) > 0$).

3. Involuntary inconvenience in the evening (16:50 - 21:00): The evening load shedding outage from 16:50 to 21:00 is the most significant period of involuntary inconvenience for households in cluster 1, 2 and 3, as shown by the red curve in Figure 6.7 (c), Figure 6.9 (c), and Figure 6.11 (c). This 4 hour and 10 minute outage includes 8 intervals of involuntary inconvenience, with seven involving multiple unschedulable appliances.

   The first interval from 16:50 to 17:10 causes 20 minutes of involuntary inconvenience. Rescheduling the appliance outside the outage adds an involuntary rescheduling inconvenience, despite reducing the loss of power supply from 0.0324 kWh, 0.0897 kWh and 0.1793 kWh to zero for clusters 1, 2 and 3, respectively. The total inconvenience J_I formulated in 3.12 includes both involuntary inconvenience and involuntary rescheduling inconvenience. In this case, this results in 40 minutes of total inconvenience.

   The second interval from 17:10 to 18:00, with three appliances that cannot be used, causes 150 minutes (2 hours and 30 minutes) of involuntary inconvenience during the 50-minute outage interval. If the household does not reschedule these 3 appliances and does not have backup power, it is inconvenienced involuntarily for 2 hours and 30 minutes and loses 0.1860 kWh, 0.4302 kWh and 0.8604 kWh of its power supply for clusters 1, 2, and 3, respectively. Should the household reschedule these three appliances to periods preceding or following load shedding outages using the load shedding forecasting model $u_{LS}(k)$ in (5.2), assuming that it does not have backup power, its power supply loss will be minimised to zero. However, this strategy introduces an additional 2 hours and 30 minutes of involuntary rescheduling inconvenience due to rescheduling appliances outside of their preferred or routine usage times, or outside times when they are needed most. This inconvenience in rescheduling is in addition to the initial 2 hours and 30 minutes of involuntary inconvenience when the appliances are not used in their preferred schedule when they are most needed. Thus, without rescheduling, the total inconvenience is 2 hours and 30 minutes. With rescheduling, the inconvenience doubles to 5 hours (300 minutes), comprising the initial 2 hours and 30 minutes plus involuntary inconvenience and an additional 2 hours and 30 minutes of rescheduling inconvenience.

© University of Pretoria

Given that the weighting factors in (5.11) are equivalent for J_I and J_{LPS}, inconvenience and loss of power supply are equally weighted. The objective function for loss of power supply enables appliance use, while the inconvenience objective function enables appliance use when it is needed most, based on regular household usage patterns. Consequently, while involuntary rescheduling minimises energy loss, it doubles total inconvenience, resulting in the same objective function value J as when appliances are not used during a load shedding outage without backup power. This observation might seem counterintuitive because one might assume that using appliances is always preferable to not using them. However, this preference holds true only if the sole concern is the usage of appliances, without considering the timing of their usage. The impact of load shedding extends beyond just energy unavailability. It also includes loss of productivity time due to appliances not being used during critical periods. Minimising only the loss of power supply ensures that appliances are used, but it is equally important to factor in inconvenience to ensure that they are used when needed most.

The third interval from 18:00 to 18:40 involves four appliances that are critical but cannot be used, resulting in 160 minutes (2 hours and 40 minutes) of involuntary inconvenience during the 40-minute load shedding outage. If these four appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 2 hours and 40 minutes, and a loss of power supply of 0.1823 kWh, 0.3627 kWh, and 0.7253 kWh, for clusters 1, 2 and 3, respectively. Involuntarily rescheduling these appliances to periods outside of load shedding ($u_{LS}(k) = 1$) eliminates power supply loss, but introduces an additional 2 hours and 40 minutes of inconvenience due to the operation of the appliances outside of their preferred usage times. Hence, without involuntary rescheduling, the total inconvenience is 2 hours and 40 minutes, while with involuntary rescheduling, the inconvenience doubles to 5 hours and 20 minutes (320 minutes).

The fourth interval between 18:40 and 19:00 involves five appliances that are important but cannot be used, resulting in 100 minutes (1 hour and 40 min) of involuntary inconvenience during the 20-minute load shedding outage period. If these five appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 1 hour and 40 minutes, and a loss of power supply of 0.1 kWh, 0.2057 kWh, and 0.4114 kWh for clusters 1, 2 and 3, respectively. Involuntarily rescheduling these five appliances to periods outside of load shedding when $u_{LS}(k) = 1$ eliminates loss of power supply, but introduces an additional 1 hour

© University of Pretoria

and 40 minutes of inconvenience due to using these appliances outside of their preferred usage times. Therefore, without involuntary rescheduling, the total inconvenience is 1 hour and 40 minutes, while with involuntary rescheduling, the inconvenience doubles to 3 hours 20 minutes (200 minutes).

The fifth interval between 19:00 to 19:10 involves four appliances that are critical but cannot be used, resulting in 40 minutes of involuntary inconvenience during the 10-minute load shedding outage. If these four appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 40 minutes and a loss of power supply of 0.0416 kWh, 0.0982 kWh and 0.1964 kWh for cluster 1, 2 and 3, respectively. Involuntarily rescheduling these four appliances to periods outside of load shedding when $u_{LS}(k)$= 1 minimises the loss of power supply to 0 kWh, but introduces an additional 40 minutes of inconvenience due to the use of these appliances outside of their preferred usage times. Therefore, without involuntary rescheduling, the total inconvenience is 40 minutes, while with involuntary rescheduling, the inconvenience doubles to 1 hour and 20 minutes (80 minutes).

The sixth interval between 19:10 and 19:40 involves five appliances that are important but cannot be used, resulting in 150 minutes (2 hours and 30 minutes) of involuntary inconvenience during the load shedding outage of 30 minutes. If these five appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 150 minutes and a loss of power supply of 0.1726 kWh, 0.4269 kWh, and 0.8537 kWh for cluster 1, 2, and 3, respectively. Involuntarily rescheduling these five appliances to periods outside of load shedding when $u_{LS}(k)$= 1 minimises the loss of power supply to 0 kWh, but introduces an additional 150 minutes of inconvenience due to the use of these appliances outside of their preferred usage times. Therefore, without involuntary rescheduling, the total inconvenience is 150 minutes, while with involuntary rescheduling, the inconvenience doubles to 5 hour (300 minutes).

The seventh interval between 19:40 and 20:10 involves four appliances that are important but cannot be used, resulting in 120 minutes (2 hours) of involuntary inconvenience during the load shedding outage of 30 minutes. If these four appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 120 minutes and a loss of power supply of 0.1726 kWh, 0.3887 kWh, and 0.7773 kWh for cluster 1, 2, and 3, respectively. Involuntarily rescheduling these four appliances to periods outside of load shedding

© University of Pretoria

when $u_{LS}(k)$= 1 minimises the loss of power supply to 0 kWh, but introduces an additional 120 minutes of inconvenience due to the use of these appliances outside of their preferred usage times. Therefore, without involuntary rescheduling, the total inconvenience is 120 minutes, while with involuntary rescheduling, the inconvenience doubles to 2 hour (240 minutes).

The eighth interval between 20:10 and 21:00 involves three appliances that are important but cannot be used, resulting in 150 minutes (2 hours and 30 minutes) of involuntary inconvenience during the load shedding outage of 50 minutes. If these three appliances are not rescheduled and the household lacks backup power, there is an involuntary inconvenience of 150 minutes and a loss of power supply of 0.2493 kWh, 0.6265 kWh, and 1.25 kWh for cluster 1, 2, and 3, respectively. Involuntarily rescheduling these three appliances to periods outside of load shedding when $u_{LS}(k)$= 1 minimises the loss of power supply to 0 kWh, but introduces an additional 150 minutes of inconvenience due to the use of these appliances outside of their preferred usage times. Therefore, without involuntary rescheduling, the total inconvenience is 150 minutes, while with involuntary rescheduling, the inconvenience doubles to 5 hour (300 minutes).

4. Minimised inconvenience during an evening load shedding outage: There are three intervals of minimised inconvenience during the load shedding outage when $u_{LS}(k) = 0$ in the evening from 16:50 to 21:00.

   The first interval of minimised inconvenience during a load shedding outage is between 17:50 and 18:00, and involves an appliance that cannot be used during the ten minute interval and is optimally rescheduled to a ten-minute period outside of the load shedding outage, with zero loss of power supply. For the same load shedding outage period, the involuntary inconvenience includes three important appliances that cannot be used during the ten minute interval, resulting in 30 minutes of involuntary inconvenience that has been reduced to ten minutes.

   The second interval of minimised inconvenience during a load shedding outage is from 18:50 to 19:00 and involves one appliance that cannot be used during the ten minute interval and is optimally rescheduled to a ten-minute period outside of the load shedding outage, with a minimised loss of power supply of zero. For the same load shedding outage period, the involuntary inconvenience includes five important appliances that cannot be used during the

© University of Pretoria

ten minute interval, resulting in 50 minutes of involuntary inconvenience that has been reduced to ten minutes.

The third interval of minimised inconvenience during a load shedding outage is from 19:10 to 19:50 and involves one appliance that cannot be used during the forty minute interval and is optimally rescheduled for a forty-minute period outside of the load shedding outage, with a minimised loss of power supply of zero. For the same load shedding outage period, the involuntary inconvenience includes four important appliances that cannot be used during the forty minute interval, resulting in 160 minutes (2 hours and 40 minutes) of involuntary inconvenience that has been reduced to forty minutes.

5. Minimised inconvenience after load shedding outage (21:40 - 22:20): The second interval of minimised inconvenience that falls outside the load shedding outage, as indicated by the secondary y-axis on the right side of the plot where $u_{LS}(k) = 1$, and the blue cluster 1,2 and 3 minimised inconvenience plots, commences at 21:40 and concludes at 22:20. The average household with low, medium and high energy consumption experiences a 40 minute inconvenience period in the evening from 21:40 to 22:20. This is because two appliances that are used during the load shedding outage are optimally rescheduled to a time after the outage, to minimise the loss of power supply, however, the rescheduled appliances' period must align closely with the average consumption baseline schedule to minimise inconvenience.

   The first optimally rescheduled appliance for cluster 1, 2 and 3 requires 0.0997 kWh, 0.2506 kWh, and 0.5012 kWh, respectively, from the grid $P_{gr}(k)$ to minimise its loss of power supply; however, it inconveniences the average household consuming low, medium and high energy for 20 minutes from 21:40 - 22:00. The second optimally rescheduled appliance requires 0.0882 kWh, 0.2442 kWh, 0.4885 kWh for cluster 1, 2 and 3, respectively, from the grid $P_{gr}(k)$ to minimise its loss of power supply; however, it inconveniences the average low, medium and high energy consumption household for 20 minutes from 22:00 - 22:20.

6. Involuntary inconvenience at night (22:50 - 00:00): The load shedding period from 22:50 to midnight involves a single continuous appliance, such as a fridge, that cannot be used, resulting in 70 minutes (1 hour and 10 minutes) of involuntary inconvenience during the load shedding outage. If the household does not have backup power, not only are they involuntarily inconvenienced for 70 minutes, they experience a loss of power supply of 0.11 kWh, 0.31 kWh and 0.63 kWh for cluster 1, 2 and 3, respectively. A continuous appliance such as a fridge cannot be rescheduled.

© University of Pretoria

CHAPTER 7 CONCLUSION

For households capable of implementing measures against load shedding, this study introduces optimal appliance scheduling by minimising the loss of power supply and inconvenience for residential homes. Among all potential strategies, it is observed that appliance scheduling with an optimally sized 9.6 kWh battery and 2.74 kWp panel array eliminates the loss of power supply probability and reduces the inconvenience by 92% when tested in South African load shedding cases. For households that do not have the financial means to invest in the recommended strategy and still want to take measures to mitigate the effects of load shedding, the KNN with a 99.6% accuracy in predicting load shedding stages can be used to reschedule appliances without any backup power, to reduce the loss of power supply by 91.7%.

In this study, in response to the severe national load shedding experienced by residents, optimal sizing and clustering strategies are used to minimise the two quantifiable impacts of load shedding for households, which are inconvenience and loss of power supply. Households are clustered according to their total daily energy consumption using K-means clustering, which creates its clusters based on the minimised variance between similar households in the same cluster and the maximised variance between dissimilar households between different clusters. The result is that the optimal number of clusters for the 781 households representing a subset of South African homes is three clusters, based on the minimum Davies-Bouldin value of 0.9 when comparing every cluster number greater than 2, namely a low-energy consumption cluster (42.28% of households), a medium-energy consumption cluster (42.0% of households) and a high-energy consumption cluster (14.72% of households). The average daily energy consumption is 4.49 kWh for low energy consumption, 10.79 kWh for medium energy consumption and 22.06 kWh for high energy consumption households. The use of clustering serves two objectives, namely: 1) to provide optimal sizing and scheduling strategies to low-energy, medium-energy, and high-energy demand households suffering from national load shedding, and 2) to

© University of Pretoria

challenge the assumption of the initial case study, which introduced a bottom-up approach to establish an appliance usage baseline from individual appliances, by introducing a top-down approach based on clustered aggregated meter readings that are disaggregated to enable scheduling from K-means clustering.

For cluster households that can implement measures to combat power outages caused by severe national load shedding, this study introduces an optimal strategy to integrate sizing and scheduling that addresses a crucial research question. What is the most effective scheduling and sizing strategy to reduce the compulsory rescheduling, inconvenience, and loss of power supply caused by load shedding, using the optimal integration of solar panels and batteries in clustered households?

This study observes that not considering scheduling in the sizing of solar panels and batteries strategy results in the solar panel array being 51.05% larger than necessary and the battery capacity being 28.56% oversized for all three clusters. Therefore, the optimal size for the solar panel array with the optimal scheduling of an average low-energy-consumption household is 0.39 kWp with an optimal 2.4 kWh battery energy storage. The optimal size for the solar panel array with the optimal scheduling of an average medium-energy-consumption household is 0.78 kWp with an optimal 4.8 kWh battery energy storage. The optimal size for the solar panel array with the optimal scheduling of an average high-energy-consumption household is 1.17 kWp with an optimal 7.2 kWh battery energy storage.

Furthermore, the total daily involuntary inconvenience for households without backup power in clusters 1, 2, and 3 is 1 130 minutes (18 hours and 50 minutes) and the total loss of power supply is 1.5268 kWh, 4.1254 kWh, and 7.4038 kWh for cluster 1, 2 and 3, respectively. Thus, the average household with low, medium, and high energy consumption loses 33.93%, 38.20% and 33.50%, respectively, of their power supply during a load shedding outage without backup power. This is the impact of load shedding on households that do not take action to minimise their inconvenience and loss of power supply. If appliances in these households are involuntarily rescheduled for periods outside of load shedding outages and periods when they are needed least, the total rescheduling inconvenience is 1 130 minutes (18 hours and 50 minutes). Although the reduced loss of power supply will only be for continuous appliances that cannot be rescheduled, the total inconvenience will be the sum of the involuntary inconvenience and the inconvenience of involuntary rescheduling.

In examining households with low energy needs, the minimised inconvenience is 150 minutes (1 hour

© University of Pretoria

and 30 minutes). Therefore, the optimal integration of optimally sized solar panels and batteries of 0.39 kWp and 2.4 kWh, respectively, for the average low-energy-consumption household of a 4.5 kWh daily demand, reduces involuntary inconvenience and forced rescheduling by 86.73% and 93.36%, respectively, and minimises the loss of power supply caused by load shedding outages to zero.

In examining households with medium energy needs, the minimised inconvenience is 150 minutes (1 hour and 30 minutes). Therefore, the optimal integration of optimally sized solar panels and batteries of 0.78 kWp and 4.8 kWh, respectively, for the average medium-energy-consumption household of a 10.8 kWh daily demand, reduces involuntary inconvenience and forced rescheduling by 86.73% and 93.36%, respectively, and minimises the loss of power supply caused by load shedding outages to zero.

In examining households with low high needs, the minimised inconvenience is 150 minutes (1 hour and 30 minutes). Therefore, the optimal integration of optimally sized solar panels and batteries of 1.17 kWp and 7.2 kWh, respectively, for the average high-energy-consumption household of a 22.1 kWh daily demand, reduces involuntary inconvenience and forced rescheduling by 86.73% and 93.36%, respectively, and minimises the loss of power supply caused by load shedding outages to zero.

© University of Pretoria

REFERENCES

[1] G. Hou, K. K. Muraleetharan, V. Panchalogaranjan, P. Moses, A. Javid, H. Al-Dakheeli, R. Bulut, R. Campos, P. S. Harvey, G. Miller, K. Boldes, and M. Narayanan, "Resilience assessment and enhancement evaluation of power distribution systems subjected to ice storms," *Reliability Engineering & System Safety*, vol. 230, Feb. 2023, Art. no. 108964.

[2] N. M. Flores, H. McBrien, V. Do, M. V. Kiang, J. Schlegelmilch, and J. A. Casey, "The 2021 texas power crisis: Distribution, duration, and disparities," *Journal of Exposure Science & Environmental Epidemiology*, vol. 33, no. 1, pp. 21–31, Aug. 2023.

[3] D. Guo, Q. Li, P. Liu, X. Shi, and J. Yu, "Power shortage and firm performance: Evidence from a Chinese city power shortage index," *Energy Economics*, vol. 119, Mar. 2023, Art. no. 106593.

[4] C. Ramaphosa. (2023). State of the nation address by the President of South Africa. [Online]. Available: https://www.stateofthenation.gov.za/assets/downloads/State-of-the-Nation-Address-2023.pdf

[5] G. Timilsina and J. Steinbuks, "Economic costs of electricity load shedding in Nepal," *Renewable and Sustainable Energy Reviews*, vol. 146, Aug. 2021, Art. no. 111112.

[6] B. B. Umar, M. N. Chisola, B. M. Mushili, C. F. Kunda-Wamuwi, D. Kafwamba, G. Membele, and E. N. Imasiku, "Load-shedding in Kitwe, Zambia: Effects and implications on household and local economies," *Development Southern Africa*, vol. 39, no. 3, pp. 354–371, Jan. 2022.

[7] D. Musademba, M. Kanyepe, L. Madiye, and T. Hove, "Effect of load shedding in Chinhoyi

© University of Pretoria

urban residential areas, Zimbabwe," *International Journal of Energy Engineering*, vol. 2, no. 5, pp. 232–241, Feb. 2012.

[8] A. Roopnarain and R. Adeleke, "Current status, hurdles and future prospects of biogas digestion technology in Africa," *Renewable and Sustainable Energy Reviews*, vol. 67, pp. 1162–1179, Jan. 2017.

[9] A. Eberhard, "The future of South African coal: Market, investment, and policy challenges," Program on Energy and Policy Development, Stanford, CA, USA, Tech. Rep. 100, Working Paper, Jan. 2011. [Online]. Available: https://www.gsb.uct.ac.za/files/sacoalstanfordpaper.pdf

[10] W. Pierce and M. Le Roux, "Statistics of utility-scale power generation in South Africa February 2023," CSIR Energy Centre, Pretoria, Gauteng, South Africa, Tech. Rep. 1, Feb. 2023. [Online]. Available: www.csir.co.za/sites/default/files/Documents/Statistics%20of%20power%20in%20SA%202022-CSIR-[FINAL].pdf

[11] R. Ngoma, A. Tambatamba, B. Oyoo, D. Mulongoti, B. Kumwenda, and H. Louie, "How households adapted their energy use during the Zambian energy crisis," *Energy for Sustainable Development*, vol. 44, pp. 125–138, Jun. 2018.

[12] P. Nduhuura, M. Garschagen, and A. Zerga, "Mapping and spatial analysis of electricity load shedding experiences: A case study of communities in Accra, Ghana," *Energies*, vol. 13, no. 17, Aug. 2020, Art. no. 4280.

[13] H. Sadeghi, D. Toghraie, M. Moazzami, M. M. Rezaei, and M. Dolatshahi, "Integrated long-term planning of conventional and renewable energy sources in Iran's off-grid networks," *Renewable Energy*, vol. 82, pp. 134–162, Jan. 2022.

[14] M. T. Nowbandegani, M. S. Nazar, M. Shafie-khah, and J. P. S. Catalão, "Demand response program integrated with electrical energy storage systems for residential consumers," *IEEE Systems Journal*, vol. 16, no. 3, pp. 4313–4324, Sep. 2022.

© University of Pretoria

[15] K. O. Akpeji, A. O. Olasoji, C. Gaunt, D. T. Oyedokun, K. O. Awodele, and K. A. Folly, "Economic impact of electricity supply interruptions in South Africa," *SAIEE Africa Research Journal*, vol. 111, no. 2, pp. 73–87, Jun. 2020.

[16] G. Hou, K. K. Muraleetharan, V. Panchalogaranjan, P. Moses, A. Javid, H. Al-Dakheeli, R. Bulut, R. Campos, P. S. Harvey, G. Miller, K. Boldes, and M. Narayanan, "Resilience assessment and enhancement evaluation of power distribution systems subjected to ice storms," *Reliability Engineering & System Safety*, vol. 230, Feb. 2023, Art. no. 108964.

[17] D. Guo, Q. Li, P. Liu, X. Shi, and J. Yu, "Power shortage and firm performance: Evidence from a Chinese city power shortage index," *Energy Economics*, vol. 119, Mar. 2023, Art. no. 106593.

[18] D. Setlhaolo, X. Xia, and J. Zhang, "Optimal scheduling of household appliances for demand response," *Electric Power Systems Research*, vol. 116, pp. 24–28, Nov. 2014.

[19] D. Setlhaolo and X. Xia, "Optimal scheduling of household appliances with a battery storage system and coordination," *Energy and Buildings*, vol. 94, pp. 61–70, May 2015.

[20] H.-T. Nguyen, U. Safder, J. Loy-Benitez, and C. Yoo, "Optimal demand side management scheduling-based bidirectional regulation of energy distribution network for multi-residential demand response with self-produced renewable energy," *Applied Energy*, vol. 322, Sep. 2022, Art. no. 119425.

[21] Y. Wang, A. O. Rousis, and G. Strbac, "A three-level planning model for optimal sizing of networked microgrids considering a trade-off between resilience and cost," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5657–5669, Nov. 2021.

[22] M. P. Bakht, Z. Salam, M. Gul, W. Anjum, M. A. Kamaruddin, N. Khan, and A. L. Bukar, "The potential role of hybrid renewable energy system for grid intermittency problem: A techno-economic optimisation and comparative analysis," *Sustainability*, vol. 14, no. 21, Oct. 2022, Art. no. 14045.

[23] L. Shen, F. Wang, F. Wang, and J. Liu, "Backup battery allocation and workload migration

© University of Pretoria

against electrical load shedding at edge," *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 16 804–16 815, Feb. 2023.

[24] B. Hedeler, H. Hellsmark, and P. Söderholm, "Policy mixes and policy feedback: Implications for green industrial growth in the Swedish biofuels industry," *Renewable and Sustainable Energy Reviews*, vol. 173, Mar. 2023, Art. no. 113098.

[25] M. Waseem, Z. Lin, S. Liu, Z. Zhang, T. Aziz, and D. Khan, "Fuzzy compromised solution-based novel home appliances scheduling and demand response with optimal dispatch of distributed energy resources," *Applied Energy*, vol. 290, May 2021, Art. no. 116761.

[26] B. Atems, J. Mette, G. Lin, and G. Madraki, "Estimating and forecasting the impact of nonrenewable energy prices on US renewable energy consumption," *Energy Policy*, vol. 173, Feb. 2023, Art. no. 113374.

[27] H. Tazvinga, B. Zhu, and X. Xia, "Optimal power flow management for distributed energy resources with batteries," *Energy Conversion and Management*, vol. 102, pp. 104–110, Sep. 2015.

[28] A. Shirsat and W. Tang, "Data-driven stochastic model predictive control for DC-coupled residential PV-storage systems," *IEEE Transactions on Energy Conversion*, vol. 36, no. 2, pp. 1435–1448, Jun. 2021.

[29] F. Yang and X. Xia, "Techno-economic and environmental optimization of a household photovoltaic-battery hybrid power system within demand side management," *Renewable Energy*, vol. 108, pp. 132–143, Aug. 2017.

[30] A. Hassan, Y. M. Al-Abdeli, M. Masek, and O. Bass, "Optimal sizing and energy scheduling of grid-supplemented solar PV systems with battery storage: Sensitivity of reliability and financial constraints," *Energy*, vol. 238, Jan. 2022, Art. no. 121780.

[31] A. A. Adeyemo and O. T. Amusan, "Modelling and multi-objective optimization of hybrid energy storage solution for photovoltaic powered off-grid net zero energy building," *Journal of*

© University of Pretoria

Energy Storage, vol. 55, Nov. 2022, Art. no. 105273.

[32] S. M. Sichilalu and X. Xia, "Optimal power dispatch of a grid tied-battery-photovoltaic system supplying heat pump water heaters," *Energy Conversion and Management*, vol. 102, pp. 81–91, Sep. 2015.

[33] Z. Wu and X. Xia, "Optimal switching renewable energy system for demand side management," *Solar Energy*, vol. 114, pp. 278–288, Apr. 2015.

[34] T. Hove and H. Tazvinga, "A techno-economic model for optimising component sizing and energy dispatch strategy for PV-diesel-battery hybrid power systems," *Journal of Energy in Southern Africa*, vol. 23, no. 4, pp. 18–28, Nov. 2012.

[35] A. M. Mahfuz-Ur-Rahman, M. R. Islam, K. M. Muttaqi, and D. Sutanto, "An effective energy management with advanced converter and control for a PV-battery storage based microgrid to improve energy resiliency," *IEEE Transactions on Industry Applications*, vol. 57, no. 6, pp. 6659–6668, Nov. 2021.

[36] P. S. Kumar, R. P. S. Chandrasena, V. Ramu, G. N. Srinivas, and K. V. S. M. Babu, "Energy management system for small scale hybrid wind solar battery based microgrid," *IEEE Access*, vol. 8, pp. 8336–8345, Jan. 2020.

[37] D. Setlhaolo, S. Sichilalu, and J. Zhang, "Residential load management in an energy hub with heat pump water heater," *Applied Energy*, vol. 208, pp. 551–560, Dec. 2017.

[38] D. Setlhaolo and X. Xia, "Combined residential demand side management strategies with coordination and economic analysis," *International Journal of Electrical Power and Energy Systems*, vol. 79, pp. 150–160, Jul. 2016.

[39] S. S. Ghayour and T. Barforoushi, "Optimal scheduling of electrical and thermal resources and appliances in a smart home under uncertainty," *Energy*, vol. 261, Dec. 2022, Art. no. 125292.

© University of Pretoria

[40] J. Yang, Q. Sun, L. Yao, Y. Liu, T. Yang, C. Chu, L. Fang, and L. Zhu, "A novel dynamic load-priority-based scheduling strategy for home energy management system," *Journal of Cleaner Production*, vol. 389, Feb. 2023, Art. no. 135978.

[41] D. H. Muhsen, H. T. Haider, Y. Al-Nidawi, and T. Khatib, "Optimal home energy demand management based multi-criteria decision making methods," *Electronics (Switzerland)*, vol. 8, no. 5, May 2019, Art. no. 524.

[42] Y. Liu, H. Li, J. Zhu, Y. Lin, and W. Lei, "Multi-objective optimal scheduling of household appliances for demand side management using a hybrid heuristic algorithm," *Energy*, vol. 262, Jan. 2023, Art. no. 125460.

[43] A. X. Wang, S. S. Chukova, and B. P. Nguyen, "Ensemble K-nearest neighbors based on centroid displacement," *Information Sciences*, vol. 629, pp. 313–323, Jun. 2023.

[44] Z. Pan, Y. Wang, and Y. Pan, "A new locally adaptive K-nearest neighbor algorithm based on discrimination class," *Knowledge-Based Systems*, vol. 204, Sep. 2020, Art. no. 106185.

[45] T. Tsoka, X. Ye, Y. Chen, D. Gong, and X. Xia, "Explainable artificial intelligence for building energy performance certificate labelling classification," *Journal of Cleaner Production*, vol. 355, Jun. 2022, Art. no. 131626.

[46] C. van Zyl, X. Ye, and R. Naidoo, "Harnessing explainable artificial intelligence for feature selection in time series energy forecasting: A comparative analysis of Grad-CAM and SHAP," *Applied Energy*, vol. 353, Part A, Jan. 2024, Art. no. 122079.

[47] K. Aas, M. Jullum, and A. Løland, "Explaining individual predictions when features are dependent: More accurate approximations to shapley values," *Artificial Intelligence*, vol. 298, Sep. 2021, Art. no. 103502.

[48] H. Tazvinga, X. Xia, and J. Zhang, "Minimum cost solution of photovoltaic-diesel-battery hybrid power systems for remote consumers," *Solar Energy*, vol. 96, pp. 292–299, Oct. 2013.

© University of Pretoria

[49] E. Sarker, M. Seyedmahmoudian, E. Jamei, B. Horan, and A. Stojcevski, "Optimal management of home loads with renewable energy integration and demand response strategy," *Energy*, vol. 210, Nov. 2020, Art. no. 118602.

[50] R. T. Marler and J. S. Arora, "The weighted sum method for multi-objective optimization: New insights," *Structural and Multidisciplinary Optimization*, vol. 41, no. 6, pp. 853–862, Dec. 2010.

[51] A. Kumar, X. He, Y. Deng, A. R. Singh, B. Sah, P. Kumar, R. C. Bansal, M. Bettayeb, and R. Rayudu, "A sustainable rural electrification based on a socio-techno-economic-environmental-political microgrid design framework," *Energy & Environmental Science*, vol. 15, no. 10, pp. 4213–4246, Jun. 2022.

[52] L. Cozzi, D. Wetzel, and B. Stéphanie, "Africa Energy Outlook 2022," International Energy Agency, Paris, Île-de-France, France, Tech. Rep. 3, Jun. 2022. [Online]. Available: https://www.iea.org/reports/africa-energy-outlook-2022

[53] A. Babajide and M. C. Brito, "Solar PV systems to eliminate or reduce the use of diesel generators at no additional cost: A case study of Lagos, Nigeria," *Renewable Energy*, vol. 172, pp. 209–218, Jul. 2021.

[54] J. T. Mensah, "Jobs! Electricity shortages and unemployment in Africa," *Journal of Development Economics*, vol. 167, Mar. 2024, Art. no. 103231.

[55] L. Vu, T. Nguyen, B. Nguyen, M. Anam, and T. Vu, "Real-time hybrid controls of energy storage and load shedding for integrated power and energy systems of ships," *Electric Power Systems Research*, vol. 229, Apr. 2024, Art. no. 110191.

[56] A. El-Zonkoly, "Application of smart grid specifications to overcome excessive load shedding in Alexandria, Egypt," *Electric Power Systems Research*, vol. 124, pp. 18–32, Jul. 2015.

[57] A. Almutairi, N. Albagami, S. Almesned, O. Alrumayh, and H. Malik, "A novel optimal framework for scheduling rooftop solar home appliances considering electricity, real pricing

© University of Pretoria

and user comfort," *Solar Energy*, vol. 262, Sep. 2023, Art. no. 111876.

[58] V. K. Maurya and S. J. Nanda, "Time-varying multi-objective smart home appliances scheduling using fuzzy adaptive dynamic SPEA2 algorithm," *Engineering Applications of Artificial Intelligence*, vol. 121, May 2023, Art. no. 105944.

[59] J. Zhou, M. Yang, Y. Zhan, and L. Xu, "Digital twin application for reinforcement learning based optimal scheduling and reliability management enhancement of systems," *Solar Energy*, vol. 252, pp. 29–38, Mar. 2023.

[60] B. Sandoval Guzmán, E. Barocio Espejo, M. Elser, P. Korba, and F. R. Segundo Sevilla, "A hybrid clustering approach for electrical load profiles considering weather conditions based on matrix-tensor decomposition," *Sustainable Energy, Grids and Networks*, vol. 38, Jun. 2024, Art. no. 101326.

[61] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, "A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs," *Applied Energy*, vol. 361, 2024, Art. no. 122943.

[62] S. Pelekis, A. Pipergias, E. Karakolis, S. Mouzakitis, F. Santori, M. Ghoreishi, and D. Askounis, "Targeted demand response for flexible energy communities using clustering techniques," *Sustainable Energy, Grids and Networks*, vol. 36, Dec. 2023, Art. no. 101134.

[63] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, "A comparative study of clustering techniques for electrical load pattern segmentation," *Renewable and Sustainable Energy Reviews*, vol. 120, Mar. 2020, Art. no. 109628.

[64] S. Twala, X. Ye, X. Xia, and L. Zhang, "Optimal integration of solar home systems and appliance scheduling for residential homes under severe national load shedding," *Journal of Automation and Intelligence*, vol. 2, no. 4, pp. 227–238, Nov. 2023.

[65] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *33rd International Conference on Machine Learning (ICML)*, M. F. Balcan and K. Q.

© University of Pretoria

Weinberger, Eds., New York, NY, USA, Jun. 2016, pp. 478–487.

[66] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[67] A. Poggiali, A. Berti, A. Bernasconi, G. M. Del Corso, and R. Guidotti, "Quantum clustering with K-Means: A hybrid approach," *Theoretical Computer Science*, vol. 992, Apr. 2024, Art. no. 114466.

[68] C. Holder, M. Middlehurst, and A. Bagnall, "A review and evaluation of elastic distance functions for time series clustering," *Knowledge and Information Systems*, vol. 66, no. 2, pp. 765–809, Feb. 2024.

[69] S. Palaniappan, S. Karuppannan, and D. Velusamy, "Categorisation of Indian residential consumers electrical energy consumption pattern using clustering and classification techniques," *Energy*, vol. 289, Feb. 2024, Art. no. 129992.

[70] H. Zhang, J. Li, J. Zhang, and Y. Dong, "Speeding up K-means clustering in high dimensions by pruning unnecessary distance computations," *Knowledge-Based Systems*, vol. 284, Jan. 2024, Art. no. 111262.

[71] D. Naik, R. Dharavath, and L. Qi, "Quantum-PSO based unsupervised clustering of users in social networks using attributes," *Cluster Computing*, vol. 27, no. 1, pp. 997–1015, Feb. 2024.

[72] W. Wu, W. Wang, X. Jia, and X. Feng, "Transformer autoencoder for K-means efficient clustering," *Engineering Applications of Artificial Intelligence*, vol. 133, Part F, Jul. 2024, Art. no. 108612.

[73] O. Dorabiala, A. Y. Aravkin, and J. N. Kutz, "Ensemble Principal Component Analysis," *IEEE Access*, vol. 12, pp. 6663–6671, Jan. 2024.

[74] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and alternative clustering methods in modern power systems," *IEEE Access*, vol. 11, pp. 119 596–119 633, Oct.

© University of Pretoria

2023.

[75] S. H. Jessen, Z. G. Ma, F. D. Wijaya, J. C. Vasquez, J. Guerrero, and B. N. Jørgensen, "Identification of natural disaster impacted electricity load profiles with K-means clustering algorithm," *Energy Informatics*, vol. 5, no. 4, pp. 59–88, Aug. 2022.

[76] R. Pan, C. Zhong, and J. Qian, "Balanced fair K-means clustering," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 5914–5923, Dec. 2024.

[77] M. Premkumar, G. Sinha, M. D. Ramasamy, S. Sahu, C. B. Subramanyam, R. Sowmya, L. Abualigah, and B. Derebew, "Augmented weighted K-means grey wolf optimiser: An enhanced metaheuristic algorithm for data clustering problems," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 5434.

[78] R. K. Ahir and B. Chakraborty, "A data-driven analytic approach for investigation of electricity demand variability for energy conservation programs," *Energy*, vol. 282, Nov. 2023, Art. no. 128939.

[79] J. T. Putra, Sarjiya, and M. I. B. Setyonegoro, "Modeling of high uncertainty photovoltaic generation in quasi dynamic power flow on distribution systems: A case study in Java Island, Indonesia," *Results in Engineering*, vol. 21, Mar. 2024, Art. no. 101747.

[80] J. Cao, R.-X. Zhang, C.-Q. Liu, Y.-B. Yang, and C.-L. Chen, "A group resident daily load forecasting method fusing self-attention mechanism based on load clustering," *Applied Sciences*, vol. 13, no. 2, Jan. 2023, Art. no. 1165.

[81] M. K. Gupta and P. Chandra, "Effects of similarity/distance metrics on K-means algorithm with respect to its applications in IoT and multimedia: A review," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37 007–37 032, Nov. 2022.

[82] J.-W. Xiao, H. Fang, and Y.-W. Wang, "Short-term residential load forecasting via pooling-ensemble model with smoothing clustering," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 7, pp. 3690–3702, Mar. 2024.

© University of Pretoria

[83] J. Li, R. Ma, M. Deng, X. Cao, X. Wang, and X. Wang, "A comparative study of clustering algorithms for intermittent heating demand considering time series," *Applied Energy*, vol. 353, Part A, Jan. 2024, Art. no. 122046.

[84] S. Mauceri, J. Sweeney, and J. McDermott, "Dissimilarity-based representations for one-class classification on time series," *Pattern Recognition*, vol. 100, Apr. 2020, Art. no. 107122.

[85] Z. Masood, R. Gantassi, and Y. Choi, "Enhancing short-term electric load forecasting for households using quantile LSTM and clustering-based probabilistic approach," *IEEE Access*, vol. 12, pp. 77 257–77 268, May 2024.

[86] A. Li, C. Fan, F. Xiao, and Z. Chen, "Distance measures in building informatics: An in-depth assessment through typical tasks in building energy management," *Energy and Buildings*, vol. 258, Mar. 2022, Art. no. 111817.

[87] A. Martins, J. Lagarto, H. Canacsinh, F. Reis, and M. G. M. S. Cardoso, "Short-term load forecasting using time series clustering," *Optimization and Engineering*, vol. 23, no. 4, pp. 2293–2314, Aug. 2022.

[88] L. Liu, X. Hu, J. Chen, R. Wu, and F. Chen, "Embedded scenario clustering for wind and photovoltaic power, and load based on multi-head self-attention," *Protection and Control of Modern Power Systems*, vol. 9, no. 1, pp. 122–132, Jan. 2024.

[89] G. Singh and J. Bedi, "A federated and transfer learning based approach for households load forecasting," *Knowledge-Based Systems*, vol. 299, Sep. 2024, Art. no. 111967.

[90] W. Toussaint and D. Moodley, "Clustering residential electricity consumption data to create archetypes that capture household behaviour in south africa," *South African Computer Journal*, vol. 32, no. 2, pp. 1–34, Dec. 2020.

[91] W. Touissant and D. Moodley, "Comparison of clustering techniques for residential load profiles in South Africa," in *Proceedings of South African Forum for Artificial Intelligence Research*, vol. 2540, Cape Town, Western Cape, South Africa, Dec. 2019, pp. 117–132.

© University of Pretoria

[92] W. Zha, Y. Ji, and C. Liang, "Short-term load forecasting method based on secondary decomposition and improved hierarchical clustering," *Results in Engineering*, vol. 22, Jun. 2024, Art. no. 101993.

[93] M. Wei, M. wen, and Y. Zhang, "A novel spatial electric load forecasting method based on LDTW and GCN," *IET Generation, Transmission & Distribution*, vol. 18, no. 3, pp. 491–505, Dec. 2024.

[94] X. Zhang, J. L. Ramírez-Mendiola, M. Li, and L. Guo, "Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study," *Applied Energy*, vol. 308, Feb. 2022, Art. no. 118335.

[95] A. Kumar and R. Mallipeddi, "A deep clustering framework for load pattern segmentation," *Sustainable Energy, Grids and Networks*, vol. 38, Jun. 2024, Art. no. 101319.

[96] M. Pullinger, E. Zapata-Webborn, J. Kilgour, S. Elam, J. Few, N. Goddard, C. Hanmer, E. McKenna, T. Oreszczyn, and L. Webb, "Capturing variation in daily energy demand profiles over time with cluster analysis in British homes (September 2019 – August 2022)," *Applied Energy*, vol. 360, Apr. 2024, Art. no. 122683.

[97] Y. Ping, H. Li, B. Hao, C. Guo, and B. Wang, "Beyond K-Means++: Towards better cluster exploration with geometrical information," *Pattern Recognition*, vol. 146, Feb. 2024, Art. no. 110036.

[98] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," Stanford InfoLab, Technical Report 2006-13, June 2006. [Online]. Available: https://ilpubs.stanford.edu:8090/778/

[99] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, "Scalable k-means++," *arXiv preprint arXiv:1203.6402*, no. 7, pp. 622–633, Aug. 2012.

[100] W. Toussaint, May 2019, "Domestic electrical load metering, hourly data 1994-2014," South African National Energy Development Initiative, University

© University of Pretoria

of Cape Town Energy Research Centre, DataFirst. [Online]. Available: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/759

[101] M. Ritchie, J. Avenant, J. Engelbrecht, A. Rix, and M. Booysen, "Optimal sizing of battery-less domestic rooftop PV systems in South Africa using a probabilistic electricity synthesiser," *Energy for Sustainable Development*, vol. 76, Oct. 2023, Art. no. 101268.

[102] M. Ritchie, J. Engelbrecht, and M. Booysen, "Predicting residential water and electricity usage profiles with a temporal histogram model," *Sustainable Cities and Society*, vol. 99, Dec. 2023, Art. no. 104884.

[103] M. Wiese and L.-M. van der Westhuizen, "Impact of planned power outages (load shedding) on consumers in developing countries: Evidence from South Africa," *Energy Policy*, vol. 187, Apr. 2024, Art. no. 114033.

[104] W. Toussaint, S. Heunis, and M. Dekenah, May 2019, "Domestic electrical load survey data 1994-2014," Eskom, University of Cape Town Energy Research Centre, Stellenbosch University, Council for Scientific and Industrial Research, DataFirst. [Online]. Available: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/755

[105] W. Toussaint, May 2019, "Domestic electrical load survey - key variables data 1994-2014," South African National Energy Development Initiative, University of Cape Town Energy Research Centre, DataFirst. [Online]. Available: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/758

[106] S. F. Koch, B. Nkuna, and Y. Ye, "Income elasticity of residential electricity consumption in rural South Africa," *Energy Economics*, vol. 131, Mar. 2024, Art. no. 107405.

[107] R. Ramadan, Q. Huang, A. S. Zalhaf, O. Bamisile, J. Li, D.-E. A. Mansour, X. Lin, and D. M. Yehia, "Energy management in residential microgrid based on non-intrusive load monitoring and Internet of Things," *Smart Cities*, vol. 7, no. 4, pp. 1907–1935, Jul. 2024.

[108] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using

© University of Pretoria

non-negative matrix factorization with sum-to-K constraint," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4430–4441, Jan. 2017.

[109] P. A. Schirmer and I. Mporas, "Non-intrusive load monitoring: A review," *IEEE Transactions on Smart Grid*, vol. 14, no. 1, pp. 769–784, Jan. 2023.

[110] M. Li, Z. Tu, J. Wang, P. Xu, and X. Wang, "Dynamic time warping optimization-based non-intrusive load monitoring for multiple household appliances," *International Journal of Electrical Power & Energy Systems*, vol. 159, Aug. 2024, Art. no. 110002.

[111] B. Liu, W. Luan, and Y. Yu, "Dynamic time warping based non-intrusive load transient identification," *Applied Energy*, vol. 195, pp. 634–645, Jun. 2017.

[112] C. Wang, Z. Wu, W. Peng, W. Liu, L. Xiong, T. Wu, L. Yu, and H. Zhang, "Adaptive modeling for non-intrusive load monitoring," *International Journal of Electrical Power & Energy Systems*, vol. 140, Sep. 2022, Art. no. 107981.

[113] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring," *Applied Energy*, vol. 185, Part 1, pp. 331–344, Jan. 2017.

[114] N. Wiroonsri, "Clustering performance analysis using a new correlation-based cluster validity index," *Pattern Recognition*, vol. 145, Jan. 2024, Art. no. 109910.

[115] L. Giannuzzo, F. D. Minuto, D. S. Schiera, and A. Lanzini, "Reconstructing hourly residential electrical load profiles for Renewable Energy Communities using non-intrusive machine learning techniques," *Energy and AI*, vol. 15, Jan. 2024, Art. no. 100329.

[116] L. Lenssen and E. Schubert, "Medoid Silhouette clustering with automatic cluster number selection," *Information Systems*, vol. 120, Feb. 2024, Art. no. 102290.

[117] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.

© University of Pretoria

[118] F. W. Watt and P. A. Campbell, "The effects of solar insolation and cloud opacity on the optimum array size for a direct-coupled solar pumping system," *Renewable Energy*, vol. 228, Jul. 2024, Art. no. 120594.

[119] U. M. Choi, "Study on effect of installation location on lifetime of PV inverter and DC-to-AC ratio," *IEEE Access*, vol. 8, pp. 86 003–86 011, May 2020.

[120] Y. Wu, Y. Zhang, Y. Li, C. Xu, S. Yang, and X. Liang, "Driving forces and typologies behind household energy consumption disparities in China: A machine learning-based approach," *Journal of Cleaner Production*, vol. 467, Aug. 2024, Art. no. 142870.

© University of Pretoria