

Data-driven approaches enabling the design of community energy systems in the Global South



Anna Clements
Lincoln College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Michaelmas 2018

Acknowledgements

My gratitude goes to my supervisor, Professor Malcolm McCulloch, for his supervision, mentorship, and guidance throughout this research. I have learnt a great deal from him. I also thank Dr Scot Wheeler for useful discussions, advice, and insightful feedback on my work and writing.

I am indebted to the communities of Lomolo B and Echariria for their welcome, friendship, and collaboration in the work of the SONG project in Kenya, which underpins this thesis. In particular, I acknowledge and thank Duncan, William, Peter, and Laban.

For mentorship, collaboration, and continued support, I thank Ed Brown, Jon Cloke, Alison Mohr, and Richard Blanchard. I am also grateful to them for introducing me to their networks of energy access researchers and practitioners and the LCEDN, and for our collaboration on the SONG project - a life-changing experience.

Thanks to Evan Kimani, who has been of great assistance these past four years on the SONG project and for my research. To Jon Leary, again for assistance in Kenya but also for not losing the filter. I also thank Jimmy Mwangi, for his time and effort on the SONG project. For reminding me to always be a pirate, I thank Arran de Moubray.

I am grateful to members of Brill Power, for their time, effort, and dedication to our collaboration, enabling me to collect important energy data. I also extend my thanks to Dr Ramon Granell, who provided invaluable guidance and expertise on clustering analysis.

Thank you to my EPG friends and colleagues. To Matt, in particular for those two weeks in Kenya in 2016; to Constance, for conversations on data analysis; to Richard, for programming assistance; to Alycia, for great editing suggestions. Thanks to my proof readers not yet mentioned: Michael Clements, Ben Dodsworth, Nina Klein, Will Clements, Tim Dodsworth, Mira Schneiders.

My gratitude goes to my examiners, Dr Dan Rogers (University of Oxford) and Prof Jatin Nathwani (University of Waterloo), for critique and suggestions that improved this work.

Thank you to the friends who have walked some or all of this crazy DPhil with me. To Nina for trail-blazing - 'may we linger in the beautiful foolishness of things'. To Mira, for laughter, cold swims, support and understanding. I am so grateful for the generous support and encouragement I have received from Juli, Chris, Neele, Lisa, David, Federico, Becky, the Wesley Mem community, and the Dodsworth-Kraft-Hauf-Rifath clan.

Thank you Ben. Grateful that we survived both theses together. Thanks to Will, for laughter and being on my side. My greatest appreciation and thanks goes to my excellent parents. To my mum, for unwavering belief in me and support. To my dad, for the same, plus academic support and very helpful feedback. Their belief in me translates to my action. My achievements are made possible by these two people. Thank you for giving me the best start.

Preface

The material presented in Chapter 3 has been published in a special issue of the IEEE Proceedings, *‘Electricity for All: Access to Electricity Issues and Solutions for Energy-disadvantaged Communities’*, as *‘Clements, Wheeler, Mohr and McCulloch: The Service Value Method for Design of Energy Access Systems in the Global South’*:

- <http://dx.doi.org/10.1109/JPROC.2019.2901215>

The electricity data set that is the output of Chapter 4 is available from Mendeley Data: *Clements and McCulloch (2019), ‘Time series electricity data for rural households using Solar Nano-grids in Kenya’, Mendeley Data, v1.:*

- <http://dx.doi.org/10.17632/4yv37hngp6.1>

Abstract

This thesis answers the primary research question:

What techniques can be developed to address the challenges of context specificity and replication at scale in the design of electricity access systems suitable for rural, low-income communities in the Global South?

This is addressed through two sub-questions:

1. *To build electricity access systems that are suitable for specific contexts, how can we take account of end-user needs, aspirations, and context in the design process?*
2. *To support replication of electricity access systems at scale, how can we use time series electricity demand data coupled with demographic data to estimate electricity demand for these communities?*

Five techniques are developed. The first is the Service Value Method, which gathers and integrates end-user needs, aspirations, and context into engineering design practice. It incorporates data-gathering methods suitable for Global South contexts.

The remaining four techniques relate to how electricity and demographic data can be used to estimate electricity demand for new communities, supporting replication of systems at scale. Time series electricity data are linked to household demographic data, and exploited to estimate a) daily electricity consumption (relevant for battery-charging systems), and b) daily load profiles (relevant for micro-grids). For each demand metric, the approach is to retain the variation in electricity use for each household across different days.

Throughout this thesis a case study from Kenya of lighting electricity data from households using two solar nano-grids is used to demonstrate these techniques. Results from the case study, comprising a dataset of household lighting electricity data, and the typical use patterns derived from them, contribute to demand estimation research for these contexts.

Contents

List of Figures	ix
List of Tables	xii
List of Symbols and Abbreviations	xiii
Glossary	xvi
1 Introduction	1
1.1 Context	2
1.1.1 The Need for Sustainable Electricity Access Systems	2
1.1.2 The Scale of the Electricity Access Challenge	3
1.1.3 Off-grid Sustainable Electricity Access Systems	4
1.1.4 Challenges in Design and Replication of Electricity Access Systems	6
1.2 Scope of this Research	10
1.3 Case Study Background	12
1.3.1 Case Study Communities	13
1.4 Research Questions and Thesis Layout	17
1.5 Key Contributions	19
2 Literature Review	21
2.1 Part 1: Gap Analysis: Designing Suitable Electricity Access Systems	22
2.1.1 Overview of Design Disciplines	22
2.1.2 Data Gathering in User-Centered and Participatory Design	24
2.1.3 Challenges for User-Centered and Participatory Design in Global South Contexts	24
2.1.4 User-Centered Design and Participatory Design Tools, Methods, and Approaches	26
2.1.5 Product-Service Systems	30
2.1.6 Summary	31
2.2 Part 2: Gap Analysis: Demand Estimation for Electricity Access Systems . .	31
2.2.1 Demand Estimation from Appliance and Time of Use Data	32
2.2.2 Demand Estimation from Measured Electricity Data	35
2.2.3 Summary	42
2.3 Part 3: Background Theory	43
2.3.1 Clustering Analysis Applied to Electricity Data	43
2.3.2 The Energy Cultures Framework	51

3	The Service Value Method for Design of Electricity Access Systems	54
3.1	Introduction	54
3.2	Service Value Method	56
3.2.1	Step 1: Data Gathering: Service Value Test (SVT)	57
3.2.2	Step 2: Data Interpretation: Service Maps	60
3.2.3	Step 3: Preliminary System Design	64
3.3	Case Study: Rural Electrification System Design in Kenya and Bangladesh	65
3.3.1	Step 1: Data Gathering: SVT Implementation	65
3.3.2	Step 2: Data Interpretation: Results	65
3.3.3	Step 3: Preliminary Design: Solar Nano-grid	68
3.3.4	Solar Nano-grid Implementation	72
3.4	Discussion and Evaluation	74
3.5	Summary	78
4	Processing Electricity Data for Rural Nano-grid	79
4.1	Introduction	79
4.2	Time Drift and Correction in Remote Sensing	87
4.3	Methodology	89
4.3.1	Definitions of Drift Rate and Time Correction	90
4.3.2	Estimating Drift Rates for Time Correction	91
4.4	Case Study	95
4.4.1	Available Data: Measured Drift Rates	95
4.4.2	Drift Rate Estimation	97
4.4.3	Time Correction	100
4.4.4	Accuracy of Time Correction	100
4.4.5	Final Result: Household Electricity Profiles	107
4.5	Discussion	108
4.5.1	Quality of Electricity Profiles	108
4.5.2	Use of the Data in Further Work	110
4.6	Summary	114
5	Estimation of Daily Electricity Consumption for Electricity Access Systems	115
5.1	Introduction	115
5.2	Method	117
5.2.1	Required Data	117
5.2.2	Data Analysis	118
5.2.3	Demand Estimation	120
5.2.4	Inferring Demand for Sizing Battery-Charging Systems	120
5.3	Case Study	123
5.3.1	Demographic Data Acquisition	123
5.3.2	Analysis: Demographic Data and Electricity use	125
5.3.3	Demand Estimation: Community X	128

5.4	Discussion	129
5.4.1	Discussion and Interpretation of Case Study Results	129
5.4.2	Sizing Key Parameters for Battery-Charging Systems	132
5.5	Summary	138
6	Typical Electricity Use Patterns of Tier 1 Users of Rural Nano-grid	139
6.1	Introduction	139
6.2	Methodology	141
6.2.1	Clustering Analysis Stages and Evaluation Criteria	141
6.2.2	Application of Clustering Analysis to Time Series Electricity Data For Each Clustering Stage	144
6.3	Case Study Results	147
6.3.1	Representative Load Profiles for Households	148
6.3.2	Communal Energy Modes Across Households	155
6.4	Discussion	161
6.5	Summary	165
7	Estimation of Load Profiles for Electricity Access Systems	167
7.1	Introduction	167
7.2	Estimation of Load Profiles from Electricity Data and Demographic Data	169
7.2.1	Method	170
7.2.2	Results for Case Study	171
7.2.3	Discussion	177
7.3	Synthesis of Measured Electricity data and End-User Aspirations	181
7.3.1	Method	182
7.3.2	Results for Case Study	183
7.3.3	Discussion	186
7.4	Summary	189
8	Conclusions and Further Work	190
8.1	Conclusions	190
8.2	Further Work	196

Appendices

A	Supporting Material for Service Value Method	201
A.1	SVT Equipment	201
A.2	SVT Implementation Recommendations	202
A.3	Summary of SVT Discussion for Kenya	203
A.4	Assumptions about Suggested Services	203

B	Additional Electricity Data Processing Steps	210
B.1	Unrecorded Time Calibrations within Monitoring Period	210
B.2	Battery RTCs were Reset to 1970	212
B.2.1	Multiple Resets	212
B.2.2	Calibration to 2016	213
B.3	Microcontrollers Logged Data with Inconsistent Frequencies	214
B.3.1	Gap-filling Algorithm	215
B.4	Addressing Logging Disruptions During Charging	217
B.4.1	Identification of Logging Disruptions	218
B.4.2	Typical Charging Profiles	218
B.4.3	Filling in Missing Charging Data	221
B.5	Power and Duration of Use Limits	221
B.6	Testing Loading Conditions to Ensure ‘Normal’ Use Data	224
B.6.1	Monthly Electricity Use Variation	225
B.6.2	Day of the Week Electricity Use Variation	227
C	Electricity Data	229
	References	237

List of Figures

1.1	Multi-tier matrices for household electricity supply and services from the Multi-tier Framework for energy access.	5
1.2	Google earth image of Lomolo B	14
1.3	Photo of Lomolo B, June 2016	15
1.4	Google earth image of Echariria	16
1.5	Thesis Layout	18
2.1	Map of design disciplines	23
2.2	Schematic of the Energy Cultures Framework	52
2.3	The Energy Cultures Framework applied to a rural village in the Global South	53
3.1	Linear representation of design phases.	55
3.2	Service Value Test method.	58
3.3	Decision tree to categorize Near-term Economic Benefit of services.	63
3.4	Service maps for household and community services for Kenya.	69
3.5	Service maps for household and community services for Bangladesh.	70
3.6	Schematic of solar nano-grid prototype and photo of Lomolo B hub.	73
4.1	Photo of household battery in the case study	81
4.2	Comparison of raw and processed electricity data using a heat map of power on the 12V battery port.	82
4.3	Flow diagram of method to infer drift rates for batteries with missing time point data.	93
4.4	Histogram of measured drift rates between time points (1) and (2) and between time points (1) and (3).	96
4.5	Plot of measured drift rates between time points (1) and (2) against measured drift rates between time points (2) and (3).	97
4.6	Heat maps for an indicative battery showing data before and after time correction.	101
4.7	Error bounds in battery time correction between time points (1) and (2).	102
4.8	Error bounds in battery time correction between time points (2) and (3) using measured drift rate.	103
4.9	95% confidence intervals on battery time correction using predicted drift rate.	105
4.10	Size of 95% confidence intervals in corrected time using predicted drift rates at time point (3).	105

4.11	95% confidence intervals on battery time correction using mean drift rate.	106
4.12	Heat map of processed electricity data for one of the households.	107
4.13	Google earth image showing Lomolo A and Lomolo B.	111
4.14	Plot of average daily electricity consumption per household against distance to the hub in minutes.	113
5.1	Method to analyze electricity and demographic data to give discrete probability density functions of daily electricity consumption for different categories of households.	118
5.2	Demand estimation tool to estimate discrete probability density function of daily electricity consumption for a new Community X.	121
5.3	Discrete probability density functions of daily electricity consumption for three indicative households	126
5.4	Discrete probability density functions and cumulative distribution functions of daily electricity consumption for households with different ranges of number of occupants.	127
5.5	Result of demand estimation tool for fictional Community X	130
5.6	Average number of times households charge per week, histogram and plot against average daily electricity consumption.	133
5.7	Schematic showing how electricity consumption is expected to change with battery size.	135
5.8	Number of batteries present at a charging hub, for 50 households with constant charging frequencies.	137
6.1	The two-stage clustering analysis for time series electricity data.	142
6.2	The elbow method to determine optimum number of clusters for k-means clustering.	146
6.3	The percentage of the variance in the data captured as the number of PCA components selected increases, for an indicative household.	148
6.4	Heat maps of days of electricity use for five clustering algorithms to demonstrate the different algorithms.	150
6.5	Heat maps and RLPs comparing results of clustering with HAC with average linkage for household 29.	151
6.6	Heat maps and RLPs showing results of clustering with HAC with average linkage for two households.	152
6.7	Heat map showing the outlier days.	153
6.8	Bar charts showing the number of outlier days for each day of the week and each month.	153
6.9	Centroids of the clusters formed when outliers are clustered using k-means, $k = 2$ solution.	154
6.10	Heat maps comparing clustering algorithms on household RLPs.	156

6.11	Plot of inertia against the number of clusters for clustering solutions using the k-means algorithm.	157
6.12	Communal energy modes resulting from clustering the household RLPs. . . .	157
6.13	Average load profile calculated across all 49 households.	158
6.14	Variance in each communal energy mode: median and interquartile range. .	158
6.15	Bar chart showing the percentage of days, for each household, represented by each communal energy mode.	159
6.16	Heat map showing the raw days of data associated with high mode.	163
7.1	Method to analyze the household communal energy modes and demographic data to obtain the communal energy modes and average load profiles for different categories of households.	171
7.2	Demand estimation tool to estimate distribution of communal energy modes and average load profile for a micro-grid feeder.	172
7.3	Distribution of communal energy modes for each household category. . . .	173
7.4	The average load profile for each number of occupants category.	174
7.5	The load profiles and variance in load profiles for each household category, derived from raw data.	175
7.6	Result of demand estimation tool for a micro-grid feeder.	176
7.7	Estimated load profile combining measured electricity data and data about aspirations for appliances.	185

List of Tables

2.1	Challenges in user-centered and participatory design in the Global South . .	25
2.2	Review of clustering algorithms	46
2.3	Linkage criteria for hierarchical agglomerative clustering	48
3.1	Template to categorize services according to three components of Ease of Delivery	64
3.2	Summary of the 7 Ease of Delivery categories	64
3.3	Services categorized by the extent electricity is a primary component. . . .	66
3.4	Services categorized according to Near-term Economic Benefit	67
3.5	Services categorized according to three components of Ease of Delivery . .	68
3.6	Services categorized according to Ease of Delivery	71
4.1	Key parameters stored by microcontrollers in household batteries	80
4.2	Summary of steps to process electricity data into household profiles	84
4.3	Time points recorded throughout solar nano-grid trial	95
4.4	Number of batteries with measured drift rates for each time period	95
4.5	Number of batteries in groups according to availability of measured drift rates	96
4.6	Batteries classified as outliers with reason for classification	97
4.7	Summary of drift rate estimates for each group of batteries	99
4.8	Measured drift rates for batteries classified as outliers	99
4.9	Time points used for time correction calculations	100
4.10	Rounding error bounds on time point readings	102
5.1	Availability of household attribute data and correlation coefficient with average daily electricity consumption	124
5.2	Number of households in each category of Number of Occupants attribute .	125
5.3	Kolmogorov-Smirnov test results for each pair of household categories . . .	128
5.4	Community X demographic data: Number of occupants	129
6.1	Number and percentage of households using each communal energy mode, and the percentage for which the mode is dominant	160
7.1	Aspirational services from Kenya SVT data, and power, time of use, and placement or use details.	184
7.2	Daily electricity consumption of additional aspirational services.	185
7.3	Estimate of end-use electricity requirement for lighting and additional services.	186

List of Symbols and Abbreviations

c	Set of centroids for a clustering solution
d	Rate of time drift
$\hat{d}$	Estimated rate of time drift
D_{cc}	Duty cycle of current controller on the battery
$D_{n,m}$	Test statistic for Kolmogorov-Smirnov test
E	Daily electricity consumption (Wh)
E_x	Household end-use electricity requirement (Wh/day)
$F(x)$	Cumulative distribution function
g	Gap between two adjacent time stamps (s)
g_m	Mode gap size between time stamps in a csv file (s)
H_0	Null hypothesis
I	Battery current on 12V port measured outside the battery (A)
I_{USB}	Battery current on USB port measured outside the battery (A)
$I_{USB,b}$	Battery current on USB port measured inside the battery (A)
I_b	Battery current on 12V port measured inside the battery (A)
k	Number of clusters
$\mathbf{L}$	A matrix of n load profiles, each of which is a vector of p average power values: $\mathbf{L} \in \mathbb{R}^{n \times p}$
$\tilde{\mathbf{L}}$	The result of the transformation of $\mathbf{L}$ when using PCA to reduce dimensionality of the input data set, $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times q}$
n	Number of samples/data points
n_d	Number of days of electricity demand data for a household
N	Number of clusters for a particular solution once outliers have been discarded
N_h	Number of households
p	Number of variables or features that comprise each data point in a data set
q	Number of principle components selected as part of dimension reduction using PCA
R^2	Coefficient of determination - the proportion of variance in the dependent variable predicted by the independent variable(s)
s	Sample standard deviation
s^2	Sample variance
S	Set of clusters
SOC	Battery state of charge
T	Time stamps recorded by the battery
t	Actual time readings
t_s	Test statistic for Student's t-test
v	Degrees of freedom

V_{out}	Voltage on 12V port of battery (V)
$\mathbf{V}$	A matrix made up of the eigenvectors of the covariance matrix of the input data in matrix form ($\mathbf{X}$), $\mathbf{V} \in \mathbb{R}^{p \times p}$
$\mathbf{V}_q$	A reduced version of matrix $\mathbf{V}$, formed by selecting q principle components, $\mathbf{V}_q \in \mathbb{R}^{p \times q}$
$\bar{x}$	Sample mean
$\mathbf{X}$	A matrix of input data, $\mathbf{X} \in \mathbb{R}^{n \times p}$, consisting of n samples and p features
$\tilde{\mathbf{X}}$	The result of the transformation of $\mathbf{X}$ by $\mathbf{V}_q$, $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times q}$
μ	Population mean of a distribution
σ^2	Population variance of a distribution
AC	Alternating current
ANOVA	Analysis of variance - statistical models used to analyze the difference in means between samples
ASCII	American Standard Code for Information Interchange
CDF	Cumulative distribution function
csv	comma-separated values
DBSCAN	Density-based spatial clustering of applications with noise
DC	Direct current
DRE	Distributed renewable energy
EAT	Eastern Africa Time
ECF	Energy Cultures Framework
EOD	Ease of delivery
ESMAP	Energy Sector Management Assistance Program
GDP	Gross domestic product
GMT	Greenwich Mean Time
GPS	Global Positioning System
HAC	Hierarchical agglomerative clustering
ICT	Information and communications technology
IDP	Internally displaced person
IEA	International Energy Agency
IEEE	Institute of Electrical and Electronics Engineers
IQR	Interquartile range
KS (test)	Kolmogorov-Smirnov (test)
LED	Light-emitting diode
MTF	Multi-tier Framework (for energy access)
NGO	Non-governmental organization
NTEB	Near-term economic benefit
PCA	Principal component analysis
PDF	Probability density function
ppm	parts per million
PSS	Product-service systems
PV	Photovoltaic
RLP	Representative load profile
RTC	Real-time clock

S.PSS	. . .	Sustainable product-service systems
SD		secure digital (memory card format)
SDG	. . .	Sustainable development goal
SHS	. . .	Solar home system
SOC	. . .	State of charge
SOM	. . .	Self-organizing maps
SONG	. .	Solar nano-grid
SVM	. . .	Service Value Method
SVT	. . .	Service Value Test
USD	. . .	United States dollar
WSN	. . .	Wireless sensor networks

Glossary

Centroids	Mean of a cluster of data points. When applied to a clustering of time series electricity data, often called a ‘representative load profile’.
Clustering analysis	Grouping a set of objects so that those in the same group are more similar to each other than to those in other groups.
Electricity access	Term used to refer to situations or systems where energy is delivered solely as electricity (which may include provision of clean cooking facilities. See ‘Energy Access’.)
Energy access	Term encompassing access to electricity and clean cooking facilities [1]. In this thesis, the term is used when referring jointly to ‘electricity access’ and the provision of energy for clean cooking not via electricity.
Global North	Term to describe the more developed regions of the world, the majority of which are in the northern hemisphere.
Global South	Term to describe less developed regions of the world, which are typically in the Southern hemisphere.
M-PESA	Mobile phone based means of money transfer.
Medoid	The closest actual data point to the centroid of a cluster.
Productive use services . .	Services that can be used for income generation.
Representative load profile	A load profile derived from a set of time series electricity data, typically by averaging over the data.
Solar nano-grid	Off-grid energy system for rural communities, smaller than a micro-grid, and with solar PV generation.
UNIX time	Number of seconds that have passed since 00:00:00 1st January 1970.

1

Introduction

Contents

1.1	Context	2
1.1.1	The Need for Sustainable Electricity Access Systems	2
1.1.2	The Scale of the Electricity Access Challenge	3
1.1.3	Off-grid Sustainable Electricity Access Systems	4
1.1.4	Challenges in Design and Replication of Electricity Access Systems	6
1.2	Scope of this Research	10
1.3	Case Study Background	12
1.3.1	Case Study Communities	13
1.4	Research Questions and Thesis Layout	17
1.5	Key Contributions	19

This thesis addresses the question:

What techniques can be developed to address the challenges of context specificity and replication at scale in the design of electricity access systems suitable for rural, low-income communities in the Global South?

There is an inherent tension in this question between ensuring electricity access systems are suitable for the local context and enabling their rapid replication to other communities. The first aspect requires a focus on the local scale of the electricity access challenge, whereas the second seeks generalization of methods, data, and learning, to achieve installations in new areas. Therefore the primary research question is answered in two parts:

1. *To build electricity access systems that are suitable for specific contexts, how can we take account of end-user needs, aspirations, and context in the design process?*

2. *To support replication of electricity access systems at scale, how can we use time series electricity demand data coupled with demographic data to estimate electricity demand for these communities?*

To give context to this research, Section 1.1 describes why electricity access systems are important, the scale of the electricity access challenge, and the range of systems that are applicable to the off-grid, rural context. The section concludes by giving context to the two specific areas in this research, covering the challenge of context specificity, and the challenge of demand estimation. Having set the scene, the scope of the research is described in Section 1.2. A case study is used throughout this thesis to demonstrate the methods and tools put forward, which is described in Section 1.3. This covers the design and implementation of two solar nano-grids in rural communities in Kenya, from which electricity demand data were monitored. The research questions answered in each chapter, and the way the thesis is structured, is outlined in Section 1.4. Finally, Section 1.5 gives the key contributions to knowledge made through this research.

1.1 Context

1.1.1 The Need for Sustainable Electricity Access Systems

In 2015, a total of 736 million people, 10% of the global population, were living in extreme poverty (classed as living on less than \$1.90 a day) [2]. The majority of these people live in the Global South: 56% were in Sub-Saharan Africa, and 29% were in South Asia [2]. In addition, most are located in rural regions: analysis of 2013 global poverty data showed that 80% of those in extreme poverty lived in rural areas [3]. There are many more people – an additional 16% of the global population – living in ‘moderate’ poverty, on less than \$3.20 a day, with a similar geographical spread and rural-urban divide [2]. The eradication of extreme poverty is the first of the sustainable development goals (SDGs) adopted by the United Nations General Assembly in 2015 [4], and considerable effort in the international development community focuses on this.

The importance of energy access (including electricity access) is recognized in SDG 7, which calls for ‘access of affordable, reliable, sustainable and modern¹ energy for all’

¹ ‘Modern’ energy refers to access to household electricity and combinations of cooking technology and fuels that are ‘clean’ - that is they do not contribute to harmful indoor air pollution. This rules out the use of unrefined biomass with traditional cooking methods, but includes advanced biomass cookstoves and biogas. [5]

by 2030 [4]. Energy access is often seen to underpin progress towards the other goals [6]. For example, access to energy can contribute to reduced poverty and increased human development through access to improved health care, education, and profitable livelihoods [7]. In one case study, it enabled increased productivity, had a positive impact on income, and improved the delivery of social and business services [8]. It can also reduce dependence on fossil fuel sources such as kerosene (used for lighting and cooking) [9], [10]; a costly fuel source that is hazardous to health and safety [11], [12].

There is an emphasis on the provision of modern energy technologies to avoid the carbon-intensive trajectory taken by developed economies since the industrial revolution; anthropogenic emissions of carbon dioxide (amongst other greenhouse gases) into the atmosphere is widely accepted as the main contributor to climate change [13]. On a local-level, residential kerosene use contributes significantly to global black carbon emissions [14] in addition to causing household air pollution. In addition, negative aspects of climate change disproportionately affect the poorest people in the poorest countries [13]. This ‘energy-poverty-climate nexus’ means that increasing energy access, mitigating climate change, and alleviating poverty, can be complementary to each other [15]. Therefore, there is a need for electricity access systems that use renewable technology to provide clean and safe energy.

However, it is worth noting that there are additional ethical considerations to be aware of in the pursuit of projects that seek to minimize greenhouse gas emissions. We argue that it is unethical for funding bodies and decision makers from high-carbon developed economies, if responsible for disseminating energy projects at scale in low-carbon countries in the Global South, to insist on project sustainability at the expense of financial viability and consequent feasibility of the venture. This hypocritical approach is an additional barrier to increasing access to electricity, thus restricting the potential for positive development impacts that electricity access can bring regardless of associated greenhouse gas emissions. Instead, and to highlight a distinct role for research and innovation in this space, effort should be made to ensure the clean and sustainable electricity access options are also the options of least-cost.

1.1.2 The Scale of the Electricity Access Challenge

According to the International Energy Agency (IEA) [16], in 2017, 992 million people did not have access to electricity. For this statistic, electricity access is defined as a household having a connection to a grid or an off-grid electricity system that can provide access to basic services

(lighting, task lighting, mobile phone charging, radio) [1]. However, this binary statistic underestimates the size of the electricity access problem, as the presence of an electricity supply does not guarantee value for an end user. Many people are under-served: there may be limited and unreliable hours of supply, poor power quality of electricity, unaffordable service, and financial or supply chain related challenges to the acquisition of appliances to use with the supply [17]. As an alternative, ESMAP (Energy Sector Management Assistant Program, a partnership between the World Bank and other development agencies) has developed the Multi-tier Framework (MTF) for energy access [18] under the Sustainable Energy for All initiative. Multiple dimensions of electricity access are taken into account as shown in Figure 1.1a: capacity, duration, reliability, quality, affordability, legality, and health impact, resulting in 6 tiers of electricity access from no access (tier 0) to full access (tier 5). The services associated with each tier are shown in Figure 1.1b. The non-governmental organization (NGO) Practical Action recommend tier 3 be used as the baseline for determining acceptable ‘energy access’, being the first tier where energy access becomes enabling [19]. As an example of the true scale of the problem, in Kinshasa, the capital of the Democratic Republic of Congo, the 2015 binary metric indicated 90% of the population had electricity access; through the perspective of the MTF, just 21% have an ‘acceptable’ level (tier 3) of electricity access [20].

Approaches to address the electricity access problem in terms of the technology deployed can be split into off-grid solutions and grid extension. Off-grid systems are electricity systems separate from the national grid infrastructure, with their own generation and distribution components. They can range from small solar products to micro-grids (described in the next section). Off-grid systems are expected to play a major role in electricity access to reach rural users where grid extension is not viable, with the IEA predicting that 70% of new rural connections are expected to come from decentralized systems [22]. The challenge is the rate of electrification in rural areas is slower than for urban areas: for example, in Sub-Saharan Africa in 2017, 67% of the urban population had access to electricity compared to just 28% of the rural population [16]. To solve the electricity access challenge, faster-paced electrification of rural areas is a necessity.

1.1.3 Off-grid Sustainable Electricity Access Systems

There are a variety of off-grid electricity access systems, ranging in the magnitude of power supplied to the end user and differing in architectures. A brief description is given here, using

Multi-tier Matrix for Access to Household Electricity Supply

		TIER 0	TIER 1	TIER 2	TIER 3	TIER 4	TIER 5	
ATTRIBUTES	1. Capacity	Power ¹		Very Low Power Min 3 W	Low Power Min 50 W	Medium Power Min 200 W	High Power Min 800 W	Very High Power Min 2 kW
		AND Daily Capacity		Min 12 Wh	Min 200 Wh	Min 1.0 kWh	Min 3.4 kWh	Min 8.2 kWh
		OR Services		Lighting of 1,000 lmhrs per day and phone charging	Electrical lighting, air circulation, television, and phone charging are possible			
	2. Duration	Hours per day		Min 4 hrs	Min 4 hrs	Min 8 hrs	Min 16 hrs	Min 23 hrs
		Hours per evening		Min 1 hrs	Min 2 hrs	Min 3 hrs	Min 4 hrs	Min 4 hrs
	3. Reliability						Max 14 disruptions per week	Max 3 disruptions per week of total duration < 2 hours
	4. Quality						Voltage problems do not affect the use of desired appliances	
	5. Affordability					Cost of a standard consumption package of 365 kWh per annum is less than 5% of household income		
6. Legality						Bill is paid to the utility, prepaid card seller, or authorized representative		
7. Health and Safety						Absence of past accidents and perception of high risk in the future		

¹ The minimum power capacity ratings in watts are indicative, particularly for Tier 1 and Tier 2, as the efficiency of end-user appliances is critical to determining the real level of capacity, and thus the type of electricity services that can be performed.

(a) Multi-tier matrix for access to household electricity supply. Table ES.2 in [21].

Multi-tier Matrix for Access to Household Electricity Services

		TIER 0	TIER 1	TIER 2	TIER 3	TIER 4	TIER 5
Tier criteria	Not applicable	Task lighting Phone charging	General lighting Television Fan (if needed)	Tier 2 AND Any medium-power appliances	Tier 3 AND Any high-power appliances	Tier 4 AND Any very high-power appliances	

(b) Multi-tier matrix for access to household electricity services. Table ES.3 in [21].

Figure 1.1: Multi-tier matrices for household electricity supply and services from the Multi-tier Framework for energy access. Source: Bhatia and Angelou [21].

the categorization in [23]. In addition, charging stations are described, a variation of which is prominent in this research. In terms of the MTF, the progression of systems laid out here is associated with increasingly higher tiers of electricity access.

Pico-solar products are the smallest systems, consisting of a small, usually sub 10W solar PV panel integrated into, or manually connected to, a lantern or other small appliance. Moving up a power level to approximately 10–1000 W, solar home systems (SHSs) are individual household systems consisting of a solar panel and battery, to power services at tiers 1–2 of the MTF (usually lighting, mobile phone charging, radio, with potential for a TV).

Charging stations enable the charging of batteries, lanterns, or mobile phones, at a centralized location [24], [25]. For the case of batteries and lanterns, the devices are charged at the hub and then carried to homes or businesses for use. Batteries may be rented, owned by end users, or a rent-to-buy scheme may be used. The centralization of the solar PV generation means that surrounding social or business ventures can also benefit from the electricity, such as in the case studies in [24] where lighting is provided to the village market. In this research, charging stations for battery charging are referred to as ‘battery-charging systems’.

Micro-grids (also known as mini-grids) have the potential to step up the power and energy provision to end users, as they are wired electricity systems that consist of generation and distribution components. Most frequently, the generation is centralized in one location, is approximately 1 kW or higher, and consists of a storage component to match supply to demand. From the centralized generation there are radial distributions lines (‘feeders’), which connect groups of households. Micro-grids can be AC or DC, providing power for basic household services (lighting and mobile-phone charging) with the potential to power higher tier services that may provide opportunity for income generation. Micro-grids are associated with productive use services more than the other systems.

1.1.4 Challenges in Design and Replication of Electricity Access Systems

1.1.4.1 The Challenge of Context Specificity

As discussed above, there is a need for sustainable electricity access systems to reach rural, low-income people in the Global South, to reduce poverty and enable a higher quality of life. However, all community infrastructure development projects in the Global South are complex to design and deliver, and there are many factors that must be considered that ensure they are a suitable solution to the context. A high-profile example of a development project often labelled as ‘failed’ is the PlayPump, which was supposed to use the movement of playing children on a roundabout to pump water from wells to give access to water [26]. A large roll-out was attempted in the 2000’s after the technology received a lot of media attention and huge donations. Incorrect assumptions were made in the use-cases in the design phase of the PlayPump, contributing to project failures when many PlayPumps were rolled-out. It was wrongly assumed that children would want to play on it, for sufficient time to acquire a usable amount of water. The actual desires and behaviours of the children were not understood or

incorporated into the design. In addition to this, cultural norms about the burden of water collection on women in some of the societies were not understood. The result of the roll-out of the PlayPump to many communities was that many PlayPumps fell to disuse. Reports circulated about women having to push the roundabout, which was less efficient than the hand-pumps that had been replaced (without asking the local community), and children being paid to push it [27], [28]. While the roll-out of PlayPumps is widely accepted to have failed and was quietly scaled back, they continue to be installed in specific contexts of large rural primary schools, where they are an appropriate technology [29]. This real-world example shows that projects must be tailored for the people and the context in which they are being installed - a one-size fits all approach does not necessarily work, and the requirement for context specificity must be acknowledged [30].

The same is true for electricity access projects. Although there is not a ‘failure’ example as famous as the PlayPump, research has shown that systems are not necessarily transferable and surrounding contextual factors matter. One example is that in the 1980’s, efforts were made to replicate the rapid market expansion of SHSs in Kenya in neighbouring Tanzania, through deployment of training that had been a catalyst in Kenya. This was not successful due to differences in the actors and context of the two countries [31]. In addition, electricity systems which do not meet the needs of the users, ignore socio-cultural factors, or fail to integrate into the surrounding context, can be met with a lack of acceptance and ownership on the part of end users, and fall into disrepair and abandonment [32].

Even if success occurs in deployment such that end users actually use the technology, whether or not electricity access systems have positive impacts, and what those impacts are, is also contingent on context. A report by Practical Action [7] found that electricity access projects are more likely to have a positive impact on poverty when the following enabling contextual factors are in place: access to finance and affordable appliances, access to markets, access to skills and knowledge.

In addition, to fully benefit from electricity access, that access must enable services that people can use to improve their quality of life and/or to generate income [17]. In particular, productive use services are crucial to the relationship between electricity access and poverty reduction [7]. Due to the need for productive use services, and the dependency on other contextual factors for positive outcomes, a service-oriented approach to electricity access is important to ensure electricity access is actually impactful. The focus is on delivery of

the whole service, rather than a supply-oriented approach that just provides power. This research adopts the service-oriented perspective of viewing electricity access as a ‘product service system’, as discussed in Chapter 2, Section 2.1.5.

This background context shows that in making design decisions such as the architecture of an electricity access system, and the services it must be able to support, a significant challenge is ensuring the system is useful to, and so used by, the end users. During design, end-user needs, socio-cultural factors and local context must be understood and integrated into the design process [33]–[35].

1.1.4.2 The Challenge of Demand Estimation

A key requirement for the design process of electricity access systems is to have knowledge of the demand for electricity. This is important for specifying the type of system, the means of electricity generation, and sizing the technical components (generation, storage, distribution) [36]–[38]. Over-estimation of demand leads to unnecessarily costly systems, barring implementation entirely or wasting money, and under-estimation has consequences for system reliability [39].

Knowledge of demand is required at different scales. Sizing components of electricity access systems placed within the household requires knowledge of demand at the household level. For shared electricity systems such as micro-grids, aggregated demand at the system level is required to size generation, storage, and distribution components. When households are aggregated, the diversity of demand becomes important. This is when diversity in the time and magnitude of demand across households means that the aggregated peak power flow is less than the sum of the individual household peak power demands [40].

Two important aspects of demand are the magnitude of electricity use per day, and the load profiles describing the temporal aspect of electricity use throughout the day. Total daily electricity consumption is required to ensure an overall balance of supply and demand in the system. The load profile is important for specifying ratings of transmission and distribution equipment, and to ensure that measures are in place so that demand is met at the time it occurs.

The aspect of demand that is required for a particular system depends on the type of system. For battery-charging systems, the generation at the charging hub and demand at the household are decoupled by the portable battery, which is physically moved between the two. Therefore, to size the portable battery, the household electricity demand per day is

required, and the load profile that gives time of electricity use is of less importance. However, for wired electricity access systems, knowing the time of use in the day is also important. The load profile is required because generation and demand are not separated by a battery that is moved between them and in the absence of storage, instantaneous demand at the household must be met by simultaneous and equal supply at the generation hub. Storage in the system may relax the requirement that supply at the generation hub must meet demand at the point of end-use, but the storage, distribution wires and associated power electronics require an estimate of the expected load profile to be sized correctly. Knowing the demand profile also informs whether demand-side response measures might be usefully incorporated in order to move demand to times of excess supply [41].

Estimating electricity demand of any form is difficult for communities that do not currently use electricity, because by nature of not having electricity, there is no pre-existing electricity demand data for that community. In general, there is a lack of demand data for rural communities in the Global South that can be used to design systems or support research and development. This is recognized by the IEEE Working Group on Sustainable Energy Systems for Developing Communities, which have an Energy Use in Off-grid Applications Task Force. In 2017 the Task Force started an archive of time series data for off-grid projects, which seeks to combat the issue of lack of data for this field [42]. In addition, situated more in the field of practitioners, the consortium of energy access stakeholders called ‘Power for All’ have recently launched the Platform for Energy Access Knowledge [43], an online tool to organize, synthesize and manage research and data for this field. There is a need for more electricity demand data to be measured and made widely available.

Further challenges in demand estimation are factoring in load growth and the effect of changes in technology. Whether or not load growth occurs, and the rate at which it occurs, is difficult to predict in this context. If expected load growth fails to occur, the consequence can be over-prediction of demand, leading to over-sizing of the electrical system, which affects the financial viability of the system. Load growth failure may be caused by a lack of ability to pay on the part of the end users, unexpected use of alternative energy sources, or a lack of end user trust in, and so use of, the system [44]. If it does occur, the rate of load growth is not easy to predict - the lack of data discussed above contributes to this issue. Changes in technology can also affect electricity demand, such as the effort for increasing the efficiency of appliances. For example, the emergence of LEDs significantly reduced

the electrical power requirements to provide lighting. The Global LEAP Awards run yearly competitions, which are helping to drive the movement towards increasing the efficiency of appliances for the off-grid energy context [45].

Without data for electricity demand, system developers commonly use energy-use surveys to estimate demand for a new community [44], [46]. Energy-use (or ‘end-use’) surveys involve asking potential customers about their current and/or aspirations for appliances. Asking them how they currently, or how they expect to, use the appliances, results in time of use and duration of use data. From this, appliance power ratings are used to estimate the electricity demand. This has been shown to be an inaccurate method to estimate electricity demand; in [44], surveys conducted by a mini-grid developer were found to over-estimate demand, with an average absolute error 400% larger than actual average consumption measured after the mini-grids were deployed (this study is discussed in greater detail in Chapter 2, Section 2.2.1.1).

Other methodologies involve using existing data from other projects, and seek to find a relationship between electricity use and descriptive data about the customer. A range of approaches in the academic literature are discussed in Chapter 2. In terms of supporting system developers to achieve replication of electricity systems in the field, there is a need for new techniques that are easily implemented and yield better estimations of demand than is provided by energy-use surveys.

1.2 Scope of this Research

The research in this thesis uses bottom-up approaches to contribute techniques to support the design and replication of off-grid, rural electricity access systems for low-income users. The tools and methods are developed to 1) ensure systems are suitable for specific contexts by integrating end-user needs and context into the design process, and 2) enable replication at scale through estimation of demand for new communities from existing electricity data and demographic data.

With respect to 1), this research presents a method to uncover and include end-user needs, aspirations, and local context in the design process by gathering quantitative and qualitative data directly from end users. It is relevant to the design of electricity infrastructure projects by multi-disciplinary teams. The scope of integration into the design process is to

the level of a preliminary design, enabling decisions about system architecture and what services it should be designed to support.

With respect to 2), the focus is on estimation of household and system level demand. To investigate the demand metrics for battery-charging systems, this research considers household daily electricity consumption. To investigate the demand metrics for wired electricity access systems such as micro-grids, the focus is load profiles that describe household electricity consumption throughout the day. These are addressed in separate chapters, as explained in Section 1.4.

The scope of the demand estimation research is to estimate these aspects of electricity demand for un-electrified communities. The design of electricity systems in terms of the sizing of technical components and system control strategy is not within the scope of this research, rather the results presented here are intended to directly feed into this next stage of the design process. In addition, detailed forecasting of future demand, or investigating how demand might change over time, is not addressed.

Throughout the thesis, the techniques presented are illustrated using a case study. This is a research project that investigated whether ‘solar nano-grids’ are viable systems for electrification of rural villages in the Global South. The project involved design and deployment of solar nano-grids, and detailed background to the case study is given in the next section (Section 1.3). The preliminary design of the system is used to illustrate the first technique presented in this thesis, addressing the inclusion of end-user needs, aspirations and context in the design process. Two systems, which include battery-charging systems to provide the household electrification, were deployed in rural communities in Kenya. Time series electricity data were collected from the battery-charging systems and are household electricity data for electric lighting (tier 1 on the MTF). These data are used to demonstrate the demand estimation techniques throughout the thesis, though the methodologies can also be applied to electricity data where there is access to more household services.

Although the thesis uses instances from the solar nano-grid project as a case study, it does not present contributions in terms of the actual design specifications of a new electricity system, in which case a quantified cost/benefit model would be important to include. Instead, it contributes methods and electricity demand data that advance design practice and seek to more accurately estimate electricity demand to enable a least-cost design of an electricity system.

Although this thesis is within the field of engineering, given the subject matter of electricity access systems, it is cognizant of social science methods to understand the socio-cultural aspects of electricity access technologies. In particular, the Energy Cultures Framework was used to understand the interacting factors that influence energy behaviour [47]. This is described in Section 2.3 of Chapter 2 and is referred to throughout the thesis when required.

1.3 Case Study Background

The SONG project (2014-2018) (EPSRC reference: EP/L002612/1) investigated whether a solar nano-grid is a technically viable, affordable and sustainable electrification method to meet community energy needs for rural villages in the Global South. The concept of a solar nano-grid sits between the functionality provided by SHSs and micro-grids, providing tier 1–2 electricity access to a small number of households (20–50) whilst powering small-scale productive use services [48]. The SONG project was a multi-disciplinary collaboration between Global North and South academics and in-country partner organizations, working with two rural villages in Kenya and two in Bangladesh to develop solar nano-grids suitable for each context.

In 2014-2015, community consultations took place in the four communities to understand current energy practices, community needs and aspirations, and the local and national energy context. From this, the architecture of the solar nano-grid was designed: a battery-charging system component, to provide the household electrification; productive use services located at the central hub, to provide income-generating opportunities for the community; solar PV generation at the hub. The household electrification component consisted of household batteries, to be charged at the hub, and a lighting system and mobile phone charger. The household batteries for the battery-charging system were built by ionQube (now Brill Power), a start-up from the University of Oxford. A prototype solar nano-grid was installed in each village in Kenya in June 2016, and a Village Energy Committee was set-up in each community to represent the communities in the project formally.

In the month after the installation, the in-county partner NGO became insolvent, and a new in-county partner had to be found. In addition, technical issues with the batteries and battery-charging system meant that batteries required frequent repairs, and the number in operation reduced throughout the trial. The monitoring period ran from June 2016 to November 2016,

when the time series electricity data were collected. Since then, development and adjustment of the systems has been on-going, led increasingly by the communities and the new in-country partner, as the system is adjusted to meet changing community needs and other external factors.

My role in the SONG project involved working on the design and installation of the solar nano-grids. In addition, my role involved project management of the development and deployment of the solar nano-grids, and I participated in designing the community consultation processes and gathering the data about the end users during the consultations. In November 2014 and January 2015, I co-led focus groups using the methodology described in Chapter 3 in Kenya and Bangladesh respectively. In July 2015 I spent a week in Kenya, participating in a second community consultation process to get feedback on the proposed solar nano-grid design. From then, I worked in Oxford on detailed design and sizing of system components, building the electrical circuits for the battery-charging systems, and co-managing sourcing of equipment for the installation. I was in Kenya for four weeks leading the installation of the systems in June 2016. In November 2016, I spent 10 days in Kenya gathering the time series electricity data from the household batteries and conducting surveys about end users' experiences.

I also spent 11 weeks living in a community in a different county in Kenya. Some of the conjecture about results in this thesis draws on observations and experiences from this time, where relevant.

The main source of data for this research comes from the Kenyan case study context. This is because the design and installation of the solar nano-grids in Bangladesh were carried out by other members of the SONG project team. Therefore, in the descriptions of the communities in the next section, more information is given on the Kenyan context.

1.3.1 Case Study Communities

1.3.1.1 Kenya

Kenya is a country in East Africa with approximately 50 million people. GDP is 1,000 USD per capita (in 2014, compared to 41,000 USD/capita for the UK) [2]. Kenya has undergone rapid electrification in recent years: in 2010, 18% of the population had access to electricity; in 2017, this had risen to 73% [16]. The rural urban divide is such that 90% of the urban population have access to electricity, compared to 68% of the rural population. There were 13 million people without electricity in 2017. The capital of Kenya is Nairobi, and the County

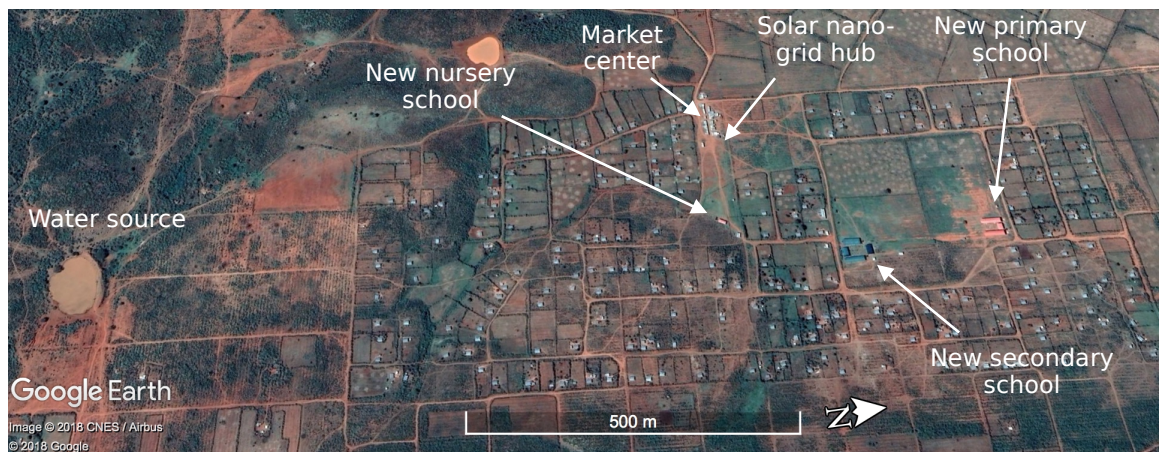


Figure 1.2: Google earth image of Lomolo B

in which the case study communities are located, Nakuru County, is 160 km North West of Nairobi. Given Kenya is on the equator, sunset and sunrise are at approximately 06:30 and 18:30 all year round, varying by up to 20 minutes either side throughout the year [49].

In Nakuru County, Kenya, Lomolo B (GPS 0°00'37.39" S 36°02'30.86" E) (Figures 1.2, 1.3) is a village of internally displaced persons (IDPs) who were moved from the Mau forest by the Kenyan government due to a conservation project and eventually settled on land in a sisal plantation in early 2014. In late 2014, when the community consultation phase of the project was carried out, Lomolo B consisted of about 250 houses, and had a small market center of several shops and a hotel (cafe/bar), a dilapidated primary school, no health center, no church building and access to water was via a dirty rain-fed pond. In July 2016, when installation of the solar nano-grid occurred, it had a new primary school, nursery school, and secondary school, and the market centre had doubled in size and was showing signs of economic growth and activity. Farm produce is sold from small kiosks around the village. It is 13 km from the nearest town, Mogotio, where most provisions come from and where health care can be accessed, reached by roads through the plantation that are poor quality.

Echariria (GPS 0°20'33.28" S 36°13'25.68" E) (Figure 1.4) is a larger, more spatially dispersed and economically diverse community of IDPs. In 2014, Echariria could be split into three areas formed by IDPs settling in stages following electoral violence in 1992, 2001, and 2007-2008, corresponding to areas 1, 2 and 3 respectively. Area 1 is easily accessible via a good road, and households are arranged relatively sparsely, at an estimated density of 210 households per square kilometer. All the main services are located in Area 1: a primary and secondary school; a main street with shops, mills, hotels, churches; an unregistered



Figure 1.3: Photo of Lomolo B, June 2016. Taken from the perspective of looking away from the solar nano-grid. Note the nursery on the right, and presence of grid poles installed in the previous month.

‘dispensary’, which is serviced by one nurse. Area 2 is less well developed, is more densely populated (approximately 1,300 households per square kilometer), and is situated up a hill on a rough road. Area 3 consisted of people living in tents in poor conditions, and was comparatively a very small area just 50m x 50m in dimension but with a very high ‘household’ density at an estimated 22,000 tents per square kilometer. Since 2014 Area 3 has dispersed as residents have moved into Areas 1 and 2 or elsewhere.

The differing household densities and geographical spread of Echariria present an interesting case for electrification. More densely populated areas are more attractive for electricity suppliers as there are more customers in the vicinity, which is particularly important for battery-charging systems given they require customers to be willing to travel physically to and from the hub. In addition, for wired electricity systems, infrastructure costs associated electricity distribution are lower if customers are close by, which will influence the affordability for customers and profitability for suppliers. Areas 2 and 3 are therefore more favourable sites to electrify from the consideration of household density. However, if density of households is inversely correlated with socioeconomic status, such that those in more dense areas are poorer than those in less dense locations (this is certainly the case for Area 3), this may reduce the number of available customers at such locations. Thus, Echariria is an example of how electrification strategies and financial viability are influenced by geographical spread of customers and socioeconomic circumstances.

In 2014 there was no grid electricity into either of the villages. However, the design and installation process occurred over nearly 2 years, and during that time the Kenyan Government was pursuing an aggressive electrification program. By mid 2015 the schools in Lomolo B and Echariria were grid-connected through Phase 1 of the Rural Electrification Authority’s

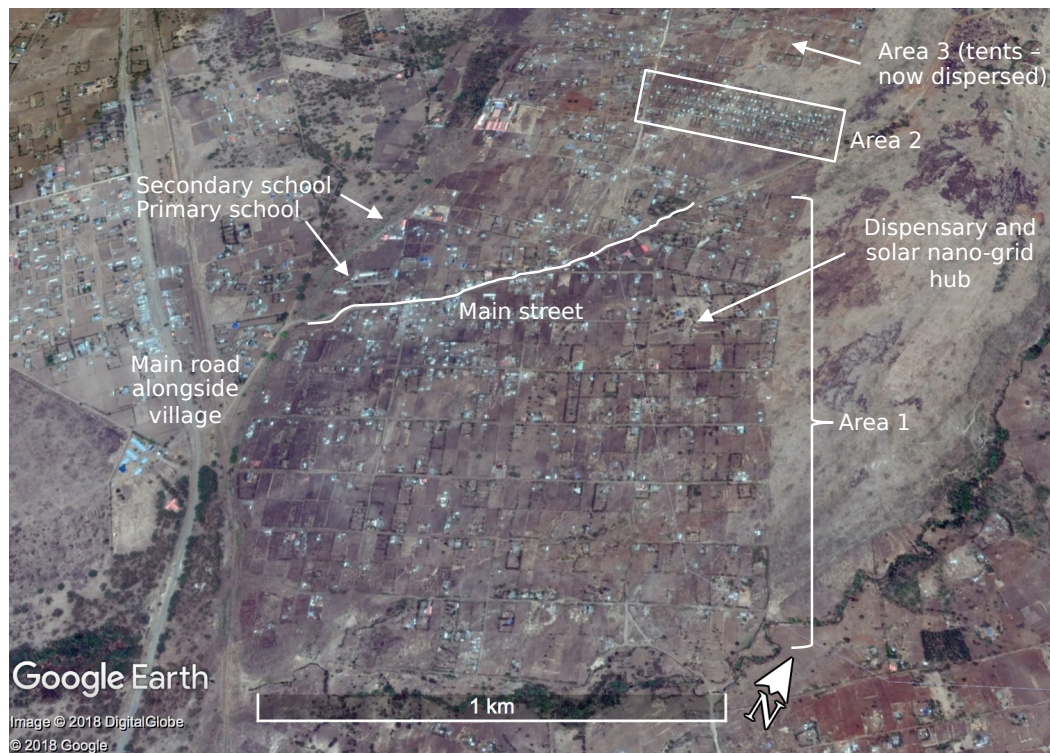


Figure 1.4: Google earth image of Echariria

strategic plan to connect all public facilities. In the first quarter of 2016, implementation of Phase 2, aiming for 65% of all Kenyan households to be grid connected by 2022, saw both villages become a target for grid-extension. Information about when this would occur was not possible to come by. In early 2016 Echariria main street was connected, and a month before installation of the solar nano-grid in June 2016 the poles and lines began to be constructed in Lomolo B (see Figure 1.3). Initially the project team was concerned that an off-grid solution as proposed by the SONG project would no longer be useful or relevant for the communities. However, through regular communications with the communities on the lead up to installation it became clear that despite the grid arrival, there was still enthusiasm for the project to go ahead. By November 2016, when user surveys were conducted, the majority of users were not connected to the grid, but nearly half were in the process of signing up.

1.3.1.2 Bangladesh

Bangladesh is a country in South Asia. It has a larger population than Kenya, with 165 million people, and also a higher percentage of population electrified at 80% in 2017 [16]. The percentage of the urban population with electricity access was 93%, and the equivalent

statistic for the rural population was 73%. This leaves 33 million people without access to electricity. It has a slightly lower GDP per capita than Kenya at 920 USD in 2014 [2].

Faitang (GPS 21°46'22.7" N 92°08'08.9" E) is a recently populated (2014) village of refugees. Approximately 50 households live in old military barracks, which are dark and poorly ventilated. Water is available through hand pumps but there are issues with high iron content, causing minor health problems. There is no health facility. In early 2015, a new primary school was in the final stages of construction, due to replace the old school. A mosque was also under construction 250 m down the access road. Faitang has good access to a nearby highway, which is also the site of a small commercial hub with a number of businesses such as mechanics, rice milling, and welding.

Baroihati (GPS 24°17'1.01" N 90°36'38.00" E) is an old (more than 100 years) village situated among rice paddy fields in Bangladesh, of approximately 25 households. It is a well-established community and in early 2015, many households had a SHS. Access to water is through several hand pumps, but there is no school or medical facility. Households are arranged in large buildings, which contain several households. The community is reasonably well connected to local villages, but to get to larger towns it is necessary to take a ferry across a river, increasing journey time.

1.4 Research Questions and Thesis Layout

This thesis examines what techniques can be developed to address the challenge of context specificity, and the need to replicate at scale, in the design of electricity access systems for rural, low-income communities in the Global South. Each chapter contributes towards answering the over-arching research question of this thesis through the two sub-questions given at the beginning of this introduction. Figure 1.5 shows the hierarchy of questions, demonstrating how the work in each chapter feeds into answering the over-arching research question.

The first sub-question addresses the need that electricity systems must meet end-user needs, aspirations, and fit into the context to be successful. Chapter 3 presents a methodology is presented to gather, interpret, and integrate these contextual data into the design process for an electricity access system. The method is demonstrated through the design of two solar nano-grids for two rural communities in Kenya.

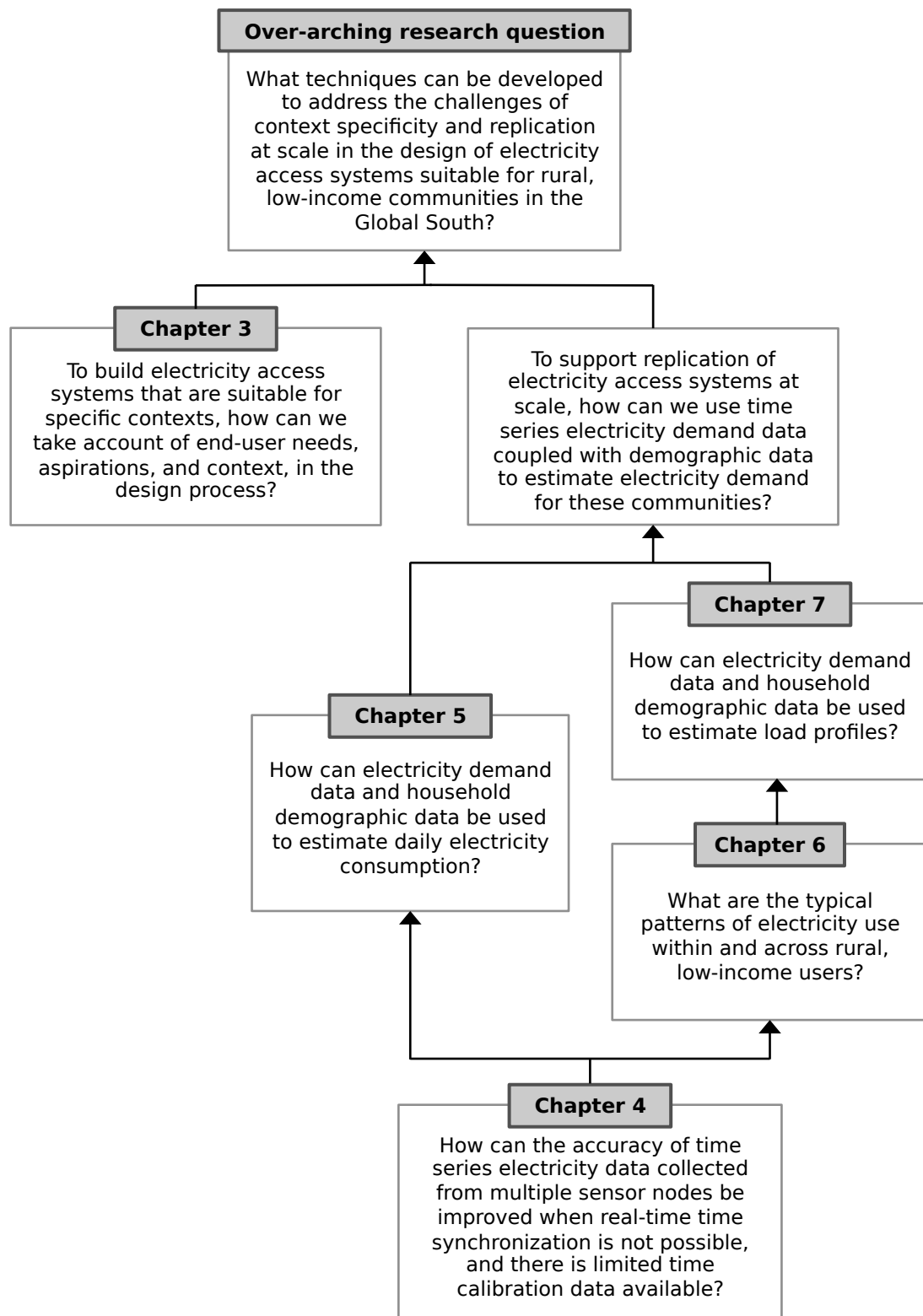


Figure 1.5: Thesis layout showing hierarchy of research questions along with the specific chapters in which they are answered.

The second sub-question addresses the challenge that to design a least-cost, reliable electricity system, an estimate of demand is required; the choice of type of system, the choice of generation means, and sizing of the system, depend on knowing the demand the system must meet. To do this, this research investigates how to use real electricity data from electrified communities to estimate demand for new projects in un-electrified communities.

Two metrics of demand are examined, each of which is important to the design of a different electricity access system. The first demand metric is total daily electricity consumption, which is required to size battery-charging systems. The second is load profiles that describe magnitude and time of electricity use, which are required to size wired electricity access systems. In Figure 1.5, the chapters where this work is presented (Chapters 5, 6 and 7) are positioned next to each other in parallel to illustrate how the research about each demand metric follows the same general steps. These steps are: 1) analysis of the data for a particular community to observe how electricity is used, 2) linking electricity use to demographic data and use of the results to estimate demand for new communities. In Chapter 5, where total daily electricity consumption is examined, both steps are presented in a single chapter. However, load profiles are a more complex and rich form of the electricity data, requiring more involved analysis to understand how electricity use patterns vary across the community. The research focus on demand estimation of load profiles is therefore split across Chapters 6 and 7.

To enable this work, electricity demand data were monitored from the household components of the solar nano-grids in Kenya using distributed sensors. A challenge in this was that there was no wireless connection between the sensors, so real-time time synchronization was not possible. Therefore, the time stamps of the time series electricity data for each household did not reflect the actual time because of drift of the clocks in the sensors, and so the data collected gave inaccurate indications of the time of use of electricity. Correction of the timestamps under these conditions is captured in the research question addressed in Chapter 4 (bottom of Figure 1.5). This supports the work outlined and demonstrated in Chapters 5, 6 and 7.

1.5 Key Contributions

The key contributions made in this research are:

1. A method to gather and interpret end-user needs, aspirations and context data, and integrate it into the design process for an electricity access system, suitable for multi-disciplinary teams and for Global South contexts (Chapter 3).
2. A method to correct the timestamps of time series electricity data for a remote multiple sensor system post data collection, when real-time time synchronization has not been possible, and where only partial data exists for sensor times and corresponding actual times (Chapter 4).
3. Electricity demand data for 51 households using battery-charging hubs in two solar nano-grids, accurate to within 5 minutes of the true time (Chapter 4).
4. Methodology to link household daily electricity consumption to demographic data, and estimate daily electricity consumption for new communities, to get key demand metrics to size battery-charging systems (Chapter 5).
5. A concise set of typical patterns of use for rural, low-income, tier 1 households in the Global South, and a process to extract these from time series electricity data, incorporating an approach to capture within-household variation (Chapter 6).
6. Methodology to link typical household electricity use patterns to demographic data, and estimate the load profile for groups of households (Chapter 7).

Contribution 1 addresses the challenge that successful electricity access systems must be appropriate for the context in which they are situated and meet the needs of the end users. Contributions 2 and 3 address the issue of a lack of publicly available electricity demand data, which are required to size electricity access systems and support research and development in this area. Contributions 4, 5 and 6 respond to the challenge of estimating electricity demand for un-electrified communities, to aid the rapid replication of systems across communities.

2

Literature Review

Contents

2.1	Part 1: Gap Analysis: Designing Suitable Electricity Access Systems .	22
2.1.1	Overview of Design Disciplines	22
2.1.2	Data Gathering in User-Centered and Participatory Design	24
2.1.3	Challenges for User-Centered and Participatory Design in Global South Contexts	24
2.1.4	User-Centered Design and Participatory Design Tools, Methods, and Approaches	26
2.1.5	Product-Service Systems	30
2.1.6	Summary	31
2.2	Part 2: Gap Analysis: Demand Estimation for Electricity Access Systems	31
2.2.1	Demand Estimation from Appliance and Time of Use Data	32
2.2.2	Demand Estimation from Measured Electricity Data	35
2.2.3	Summary	42
2.3	Part 3: Background Theory	43
2.3.1	Clustering Analysis Applied to Electricity Data	43
2.3.2	The Energy Cultures Framework	51

This literature review performs two functions. The first is a review and gap analysis of each of the topics of research associated with each sub-question. Part 1 reviews how end-user needs, aspirations, and context are included in the design process to design suitable electricity access systems. Part 2 reviews demand estimation for electricity access systems. The second function is to describe background theory relevant to the methods of analysis used in this research. The first section of Part 3 reviews the theory of clustering analysis applied to electricity data (a particular aspect of clustering analysis, the strategy used, is introduced and

reviewed in Part 2, because it is relevant to the demand estimation review). The second section of Part 3 reviews the Energy Cultures Framework, a tool to understand energy behaviour.

2.1 Part 1: Gap Analysis: Designing Suitable Electricity Access Systems

2.1.1 Overview of Design Disciplines

Design theory has a range of disciplines that describe the approach to including end-user needs and context in the design process. The two over-arching design disciplines concerned with inclusion of the end user are user-centered design and participatory design (sometimes more recently known as co-design). There are slightly different interpretations of what each involves [50], but a significant difference between them is in the role of the end user [51]. In user-centered design, the end user is a subject: they are asked for information and feedback, but are not in control of design decisions. In contrast, in participatory design the end user is a partner in the design process with design decisions taken jointly, or perhaps wholly, by end users. For example, design concepts, requirements, or use cases may be generated or developed in mixed groups of designers and end users, or end users may be enabled by participatory tools developed for the purpose to generate concept ideas more autonomously, which the designer then helps to give form to [51]. The role of the researcher or designer shifts towards facilitation.

Stemming from user-centered design and participatory design, various design disciplines have emerged since the 1970s in response to shifting perspectives about what design can be applied to. The boundaries of what can be designed have broadened beyond products, to include interactions, services, social problems and organizations [51]–[55]. This is shown in Figure 2.1, which presents design disciplines in approximate chronological order of emergence, and the change to what is being designed on a categorical y-axis. User-centered design and participatory design blur together after the initial definition to reflect the vagueness of specific tool and method attribution, and the fact that later disciplines use tools from both. It should be noted that this is one interpretation of design theory evolution, picking out some of the main trends and emerging disciplines, and a more in depth account can be found in [56].

The evolution in design disciplines reflects the increased understanding that a certain product or technology is part of a wider system, which includes the end users, their interactions

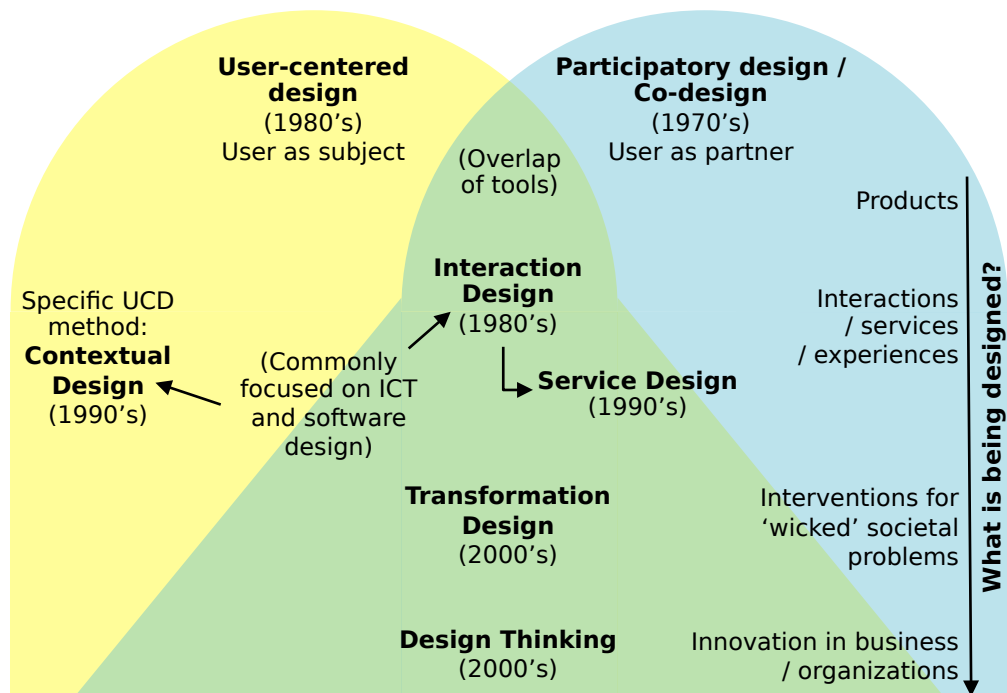


Figure 2.1: Map of design disciplines, with decade of emergence and indication of what is being designed along the y-axis.

with it, and the wider problem the technology is there to address. It also indicates increased awareness of the importance of understanding the end user and the context.

Transformation design is arguably the most relevant to the electricity access challenge and the design of electricity access systems. It was coined in a White Paper by Cottam et al. [54] from the UK Design Council (independent charity and advisor to the UK government on design), and developed outside of traditional academia ‘to bring about practical solutions to familiar and intractable social and economic problems’. This picked up on discussions dating to the 90’s regarding design problems being ‘wicked’ [57]. ‘Wicked’ problems are indeterminate, ill-formulated, involve many stakeholders with conflicting values and requirements, have numerous variables and are evolving [58]. Sustainable electricity access systems for poverty reduction in the Global South are a strong fit given their complexity, their impact on poverty, and the additional issue of sustainability. Some of the challenges in transformation design are noted as a lack of tools and techniques for multidisciplinary collaboration and the lack of a shared language to communicate between disciplines. There is a need for tools and methods that enable multidisciplinary groups, like those required to solve complex challenges such as electricity access, to work together.

2.1.2 Data Gathering in User-Centered and Participatory Design

In both user-centered and participatory design approaches, the aim is to uncover and include data about the end users in the design process, to produce more suitable products, services, or systems. The data that are of interest include socio-cultural factors about the end user (their needs, aspirations, values, beliefs), end-user practices (what they do, how they do it), and the context in which they live (factors external to the user that affect their environment and lives; they might be political, economic, social, surrounding infrastructure and services). This can also be framed using the Energy Cultures Framework (see Section 2.3.2), considering the material culture, practices, norms, and external factors that contribute to energy behaviour.

These design disciplines were largely developed outside of the rural Global South context, and require adaptation for application there given the circumstances of those settings.

2.1.3 Challenges for User-Centered and Participatory Design in Global South Contexts

There are six important challenges associated with applying user-centered design and participatory design techniques of both qualitative and quantitative natures in a Global South context, particularly the rural and low-income context of off-grid electricity systems. Some of the challenges relate to the capacity or situation of end users and stakeholders. However, (Global North) design teams that work in Global South contexts can bring with them a lack of knowledge and limited perspectives that should also be acknowledged. In Mohr [66], it is argued that Global North actors working in electricity access in the Global South should give local knowledge and understandings equal weight with their own expertise. Cloke et al. [67] also emphasize that energy ‘literacy’ should be treated symmetrically across all project partners, and rather than the focus of capacity building being on the end users, it should be equally important for other partners (designers, stakeholders). In this research, challenges for user-centered design and participatory design are examined from the perspective of community participants. However, the limitations of Global North researchers and designers in a project are an important consideration, and further discussion is found in [66], [67] .

The challenges listed in Table 2.1 have been gathered from literature on user-centered design and participatory design applied to Global South contexts, ranging from the design

Table 2.1: Challenges in user-centered and participatory design in the Global South

Challenge	Consequence
Participants: lack of understanding of design process	Hard to engage the users in user-centered design exercises, and hard to recruit them [59]. Difficult to get feedback on prototypes because participants did not understand that they represented an end-product [34], [59], [60]. Low fidelity prototypes too abstract, causing confusion and misunderstanding [34], [61]. Participants lacked the capacity to participate and be creative in a participatory design process - 'true co-design' not possible [62].
Participants: lack of understanding of technology	Users were intimidated by the prototypes [63]. Difficult to get input on ideas or feedback on prototypes [34], [64].
Participants: low literacy	Questionnaires did not work - incomplete or unanswered [59].
Language barrier between participants and designers	Gathering feedback was difficult [59]. Working through a translator had challenges in terms of clarity of questions and understanding answers [62].
Participants are in remote/ different locations to designers	Poor transport infrastructure, lack of electricity, and safety concerns limited time on site [64]. Limited money and time meant designers could not get to end users to gather data [65]. end users were in poor rural areas, other stakeholders were in a city with full-time job commitments, meaning all participants could not be brought together for central workshops, and time for participation was short [62]. Building relationships and trust between designers and participants requires considerable time investment and physical presence, and introduction/acceptance by local authority or trusted figures [62], [63].
Cultural norms and beliefs can limit participant feedback, and be a barrier to participation	Users did not want to be impolite so did not give useful criticism of prototypes [64]. High fidelity prototypes were not successful as people only offered compliments and appraisals, not wanting to cause offense by criticizing a costly object [59]. Hierarchical social structures meant it was difficult for participants of different social standing to collaborate and contribute equally when in groups together [62]. Participants were unwilling to show ingratitude by criticizing ideas or prototypes due to Buddhist beliefs [62]. Children (end users) were unwilling to put forward opinions to open-ended questions as culturally they are taught not to [62]. Cultural norms meant it was not possible to bring children (end users) to central workshops, or meet outside of their homes, as their parents were worried about child abuse [62]. Children expected to help with chores at home, so could not spend much time on the project [62].

of ICT, to medical kits and micro-finance systems. Table 2.1 also gives examples of how each challenge has manifested in previous research and design projects.

User-centered and participatory design tools and methods to uncover and include end-user needs and context in design processes must be suitable to work within the special circumstances of the Global South. The following sub-section reviews common user-centered and participatory design tools, their applicability to design for the Global South, and how in some cases typical techniques have been adapted to be more appropriate.

2.1.4 User-Centered Design and Participatory Design Tools, Methods, and Approaches

User-centered design and participatory design tools and methods can be split into two groups: some are concerned with the gathering of end-user data, and some are focused on the integration of end-user data into the design process. The tools that gather end-user data can be quantitative or qualitative. Quantitative tools gather numerical data that can be used in statistics about end users. Qualitative tools gather non-numerical data, uncovering reasons and opinions often through descriptive phrases or narratives. Employing both quantitative and qualitative methods is useful to obtain both numerical data and deeper insight and understanding of end-user responses.

A comprehensive collection of tools focused on gathering data from the end user at all stages of the design process has been collated by IDEO, a design consultancy, in their toolkit of design tools for social innovation [68]. Examples of methods are: interviews with participants singly and in group formats, spending time immersed in the end-user context, preparing conversation starters to get participants into discussion, prototyping, and asking participants to sort cards in order of importance. IDEO also outlines several pictorial methods, such as using picture cards in a card sort or asking participants to draw, which help to cope with language barriers or low literacy levels. Another successful example of using pictorial methods to engage participants in rural villages in the Global South is Hirmer's 'User-Perceived Value Game' [69]. Utilized in Uganda, the aim was to understand end-user values to more appropriately market development initiatives to increase end-user buy-in and acceptance. The game involved asking users to select and rank cards with familiar, every-day objects on them, and then explain the importance of them. This provided dual

quantitative (the ranking) and qualitative (the discussion) data that were used to characterize end-user values and reasons behind them.

Two common methods in user-centered design to help designers integrate end users into the design process are personas and scenarios. A persona represents typical users, and is a description of who they are, their behaviour, their needs, and attitudes [70]. Personas give a design team a shared understanding of the target user. A scenario is a sequence of events, such as a typical day, often constructed around the personas devised previously, that helps the designer to give context to the persona [71]. However, both personas and scenarios rely on designer assumptions to construct or else require lengthy ethnographic research that may not be possible in a rural Global South context.

Faced with the challenge of accessing participants to gather data to create personas and scenarios, Putnam et al. [65] outline a method to mine existing data for this purpose. The aim is to reduce the need to gather new data for each context when time and financial barriers exist. They use data collected in a previous research project through surveys and focus groups, about general attitudes and behaviours about technology use in Kyrgyzstan, to design a business directory app for mobile phones. This approach relies on suitable data existing for the local context and being accessible, which is not common for rural Global South contexts (see Section 2.2).

The usefulness of personas has also been criticized in terms of the efficacy with which they represent large user groups well given the wide diversity and unpredictability of human behaviour [72]. A useful contribution from Beck et al. [73] describes a participatory card-sorting method that involves end users and designers working together to create use scenarios and early design ideas, improving the accuracy of what is developed. In general, while personas and scenarios can be useful tools to visualize and consider end users in the design process, they are likely best suited to contexts familiar to the designer.

A method that can be used to understand participants' perceptions of a particular subject is free listing [74]. Free listing is a structured interviewing method from anthropology that asks participants to list everything they can think of about a particular subject. It enables the researcher to understand the scope and boundaries of a certain domain, and when repeated across numerous participants, items mentioned more frequently can be assumed to be more important or relevant. It takes the important approach of allowing participants to suggest words without influence or suggestions from facilitators. However, it does not

give participants a chance to explicitly assign relative importance to their responses, as they may not list things in order of importance.

As well as the aforementioned single-technique tools for gathering end-user data or using that data in the design process, there are also methods that address the whole design process within the user-centered design approach. Contextual design utilizes a data collection method called contextual inquiry, where the designer spends time with the end user while they carry out particular actions in their life or at work, conversing and observing how they do particular tasks [75]. It is based on the premise that simply asking someone how they do something will not get accurate responses because people are not good at observing and accurately reporting on themselves. After contextual inquiry, contextual design consists of a series of data interpretation and consolidation steps. Contextual design is a strong example of a structured methodology that details how to gather end-user data and use it in the design process. However, there may be challenges in performing contextual inquiry for electricity access projects. There may be cultural barriers, language barriers, and site accessibility issues that prevent designers spending time with end users while they go about their daily lives while conversing with them. In addition, contextual inquiry relies on confident discussions about the norms and practices of end users, which may not be in the skill-set of an engineer.

Recognizing that those unfamiliar with social sciences require specific tools to enable them to take account of socio-cultural issues in the design process, Pereria et al. [76] put forward a ‘value-oriented and culturally informed approach’ (VCIA) for the design of interactive systems. It includes ‘artifacts’ (tools) that encourage the designer to consider values and culture explicitly throughout the design process, for example through value comparison tables and a ‘culturally aware requirements table’. They apply the methods and approach to design of a social network for Brazilian teachers working with children who have a disability, and find the VCIA toolkit useful to draw attention to cultural issues designers might otherwise overlook. They note the difficulty in building tools that are easily used by those without social science experience, but which don’t oversimplify complex concepts such as culture. The VCIA provides a useful structure for progressing the design with values and culture in mind, with a framework suitable for use by engineers, but has less of a structured approach towards inclusion of surrounding context, which is equally important.

Prototypes are used in many design projects to elicit feedback from participants, often deployed periodically during an iterative design approach. However, there are problems

reported using prototypes in Global South contexts, such as participant misinterpretation and confusion (Table 2.1). One method to address these is with ‘technology artifacts’: off-the-shelf items that are functional and easily understood by participants without having to consider abstract design concepts [34], [60]. They allow for open dialogue between designers and end users about functionality and end-user experiences with the item, and circumvent the issue of participants not understanding the design process or new technologies.

To address the challenge of limited participant understanding of the design process or technology, Maunder et al. [34] and Hussain et al. [62] advocate for an approach that builds local capacity for participation. This ‘co-evolutionary approach’ requires considerable extra time and resources, training and preparing the users and their environment for the design process, but enables more equal and full participation in the design process by end users. Hussain et al. [62] notes that over time it was possible to move towards more equal participation, as participants became more psychologically empowered through training and experience. This approach expands the role of design projects in such contexts to include local capacity building of participants, and aims to move towards a sustainable situation where external designers are no longer needed.

Watkins et al. [59], Chetty et al. [64], and Hussain et al. [62] note the importance of adapting tools, methods, and design approaches according to what is appropriate in terms of the cultural norms of the participants. This requires contextual research by the design team and/or collaboration with partners familiar with the participant context in designing tools and exercises to use with participants. Chetty et al. [64] note that design for contexts with which one is not familiar requires this extra effort to incorporate participants’ values into the processes created, as well as into the system being designed. This approach ensures methods are designed so as to be suitable to enable participation of end users, taking account of cultural norms (the 6th challenge in Table 2.1). It can also mitigate four other of the challenges listed in Table 2.1: limited understanding of technology, the design process, and literacy and language barriers.

Finally, Maunder et al. [34], reviewing interactive systems design for the developing world, notes that there are no user-centered design tools that do enough to uncover complex socio-cultural issues early on in the design process to avoid failures or costly changes later. They state that the practice of ethnographic techniques is the only option currently, but these are not integrated into the design process or necessarily appropriate for non social scientists

to carry out. They also call for a way to include broader social and physical factors that might influence design decisions in the early stages. Similarly, in a paper detailing a case study addressing energy needs for a rural village in Mali, Bryden et al. [35] calls for a ‘structured way of gathering information’ to guide formation of the problem definition stage in an engineering design process. They argue that without this, engineers are at risk of making assumptions about end users and context that ultimately lead to project failure.

2.1.5 Product-Service Systems

A useful context originating outside of design theory and methods is the concept of product-service systems (PSS). The PSS model shifts away from traditional economic models of selling a product (e.g.: air conditioning unit), towards selling a combination of products and services (e.g.: thermal comfort) that fulfils a particular demand [77]. Therefore, the economic value to the consumer can be decoupled from material and energy consumption, so the provider is incentivized to reduce resource use as much as possible [78]. This was extended in the 1990’s to the concept of Sustainable Product-Service Systems (S.PSS), which are PSS offer models which offer environmentally and ethically beneficial solutions [79]. The PSS/S.PSS perspective means that the ‘offering’ being sold/bought is inclusive of all products and services that enable the production of the function, or outcome.

Research on the application of S.PSS to distributed renewable energy (DRE) has noted that combining S.PSS and DRE is a ‘win-win’ approach to addressing the energy access challenge because of economic, environmental, and socio-ethical benefits [78], [80]. For example: low-income users might be able to access services a traditional product supply model would prohibit due to the expense of the equipment, and producers are incentivized to minimize resource use, producing a capital-efficient system, the lower costs for which are passed on to the customer. In addition, a new design approach called ‘sustainable design for sustainable energy for all’ details the design process of S.PSS DRE systems [80]. There are a number of tools within the methodology that guide designers to consider all aspects of S.PSS when developing an S.PSS energy system, from training services to payment mechanisms, using templates, prompting cards, and other processes. The methods are about the process of design, organizing data, ensuring the right questions are given attention to, rather than being tools to gather the data required.

2.1.6 Summary

There are numerous calls in the literature for tools and methods that uncover socio-cultural and contextual factors, so that designers can take account of them to design successful systems. These tools and methods must be suitable for multi-disciplinary teams, because complex social problems such as electricity access require a multi-disciplinary approach. For example, they must be suitable for engineers, who are not typically trained in social-science ethnographic methods.

They must also be appropriate for the low-income, rural contexts of the Global South. User-centered and participatory design methods that interact with end users must reduce reliance on end users having a good understanding of the design process. Talking about what end users know, rather than introducing technology that they may find intimidating, or be unable to offer opinions on, is also important. Participation must be enabled at a level suitable for end-user capacity. In addition, methods must be appropriate for literacy and language barriers - pictorial methods have been successful in achieving this. Although it is desirable to spend as much time as possible with end users to understand their needs and context, in reality, money, time, and remoteness constraints are likely to limit this. Therefore, methods to interact with end users must be effective in terms of information gathered for the time taken, given it will not be possible to do a full ethnographic study for all villages requiring electricity systems. Participants must not be influenced by assumptions from the design teams during the data-gathering stage, and data must be robustly integrated into the design process so designers' perspectives do not re-interpret data inaccurately. Finally, methods must be culturally appropriate and implemented adhering to cultural norms. For this, co-design of methods alongside those familiar with end-user context is an important approach, as well as building in method flexibility so it can be adapted to different cultures and end-user capacities.

2.2 Part 2: Gap Analysis: Demand Estimation for Electricity Access Systems

As discussed in the introduction, demand estimation is crucial for the design of electricity systems in terms of specifying architecture, generation method, and component sizes, to minimize cost while ensuring a reliable system [39]. This section reviews the field of

electricity demand estimation for households and electricity access systems in Global South rural villages, drawing on methods used for other contexts where appropriate.

2.2.1 Demand Estimation from Appliance and Time of Use Data

A common method of demand estimation is a bottom-up strategy, where knowledge of appliances in the household, business, or other entity, along with use patterns (time and duration of use), is used to build up the demand load profile(s). There are different ways of obtaining the appliance and use pattern data. One method is to use energy-use (or end-use) surveys directly in the targeted village, which involves asking respondents to estimate or predict their current and aspirational energy/electricity consumption [44]. The surveys vary in what information is gathered, and how that information is used. Micro-grid developers use energy-use surveys to ask about current and aspirational appliance use, to scope out a potential site to ascertain if there is sufficient demand to be economically viable, and to estimate the load profile [44], [46]. By including questions about what new appliances end users hope to acquire to use with the electricity, these surveys attempt to take account of how electrical load might grow in the future given the appliances end users are expecting to attain. In academic literature, energy-use surveys have been used with a focus on asking about current energy use, rather than aspirational. For example, in Sandwell et al. [81], energy-use surveys are performed in un-electrified villages to increase baseline knowledge of current energy consumption and activity patterns of end users. In the same study, the data are used to infer household use profiles of lighting and mobile phone charging, which are then used to estimate electricity demand at the system level for a hypothetical village.

If data are not gathered directly from the target village through energy-use surveys, appliance ownership and use patterns are themselves estimated from other data sources or via other methods. In Blum et al., [82] appliance and use pattern data for a typical village in Indonesia are obtained from existing reports and carefully referenced literature, informed by the authors' field experience. To estimate appliance ownership and use patterns for an off-grid system in India, Sen et al. [38] uses government records for similar rural areas, literature (un-referenced) and 'expert opinions and personal judgments' - although it is not clear which numbers come from which source. Sandwell et al. [81], having used energy-use surveys to construct a basic load profile for un-electrified Indian villages, draws on other literature and census survey data to estimate an aspirational load profile. The sources are

carefully referenced, but strongly draw on un-referenced data from Sen et al. [38]. Fabini et al. [83] puts forward an approach to estimate induced residential electricity demand, by estimating appliance ownership for wards (administrative divisions) in Kenya according to socioeconomic similarity. For areas of the country that do not have appliance ownership and electricity access data from government surveys, these data are estimated according to the similarity of the area to other wards with existing data. This gives high-level indications of percentage ownership of appliances, from which electricity demand is estimated using assumptions about use patterns from other literature.

Compared to using energy-use surveys specific to the village for which demand needs to be estimated, there are several issues with these approaches. Using data sources that give aggregated data information for large regions, such as a census or government surveys, assumes that this is applicable for a particular village in question. These data are usually presented as averages, whether it is average household size, or appliance use, and the approach assumes the particular village aligns with this average, which may not be the case. In addition, these data are not always available, as shown by efforts by Fabini et al. [83] to estimate it where it is missing. With regards to use of previous literature for appliance specification data, great care must be taken that it is relevant and up to date, particularly given technological developments in electricity access - some previous literature being drawn on for appliance data was 9 years older than the study in question in [82]. Where it is not referenced, it is not possible to verify the accuracy of the data, and this is problematic if it is then used as the basis for further studies. Similarly, where assumptions made by personal judgment, or experts, are not detailed, it is not possible to judge their accuracy or adjust them to new contexts. Qualitative information directly from the field has the potential to ensure accurate adjustments of other data for the context, making it more appropriate, but an element of subjectivity is involved as it relies on external people's perceptions.

Once the appliance and use pattern data are acquired or estimated, bottom-up approaches to estimate individual customer demand or the whole system demand are used. The simplest of these is to sum the products of the appliances, power ratings, and duration of use, for each customer, to give either a daily electricity demand or a load profile if time of use is retained for each appliance. This is demonstrated in [44]. More sophisticated methods to estimate load profiles use simulation techniques to model the variation in time of use across days and households using stochastic methods. Boait et al. [40] presents a tool to estimate aggregated

demand accounting for demand diversity, from inputs of appliances, their power ratings, and probability of use, outputting the average, peak, and standard deviation in power expected for that time point. Mandelli et al. [84] puts forward a bottom-up method to formulate a load profile using a stochastic approach that introduces parameters that describe the variance in time and duration of use. The load profiles for each appliance are derived (using the variance parameters), and are then aggregated with profiles for other appliances, and other users, to give the final profile. Narayan et al. [85] uses a somewhat similar bottom-up approach, focusing on forming typical load profiles for each tier of the MTF using the latest efficient DC appliances.

However, all of these approaches rely on appliance and time of use input data, and thus their accuracy depends on the accuracy of that information. In whatever way the input data are obtained, whether through energy-use surveys, or taken from government surveys or census, the methods rely on respondents to accurately recall or estimate their own current or aspirational appliance use. As now discussed, there are significant issues with the accuracy of data gathering processes that rely on respondents to estimate or predict their behaviour.

2.2.1.1 Accuracy of Energy-use Surveys

In Blodgett et al. [44], the accuracy of energy-use surveys in predicting electricity demand for eight rural mini-grids in Kenya is investigated. This study is important because it considers a real-world situation, analyzing the error in the demand predicted from energy-use surveys actually carried out by a developer for the eight sites before installation of the mini-grids, compared to the measured electricity use once the systems were installed. The energy-use surveys conducted by the developer just before mini-grid installation asked about current appliances and associated use patterns, and also what appliances customers hoped to acquire to use with the electricity, and how they hoped to use them. However, there was no time-frame specified for this question. Using the original survey results, the authors estimate average daily consumption for 176 customers, using a simple summation of the power rating and duration of use for each appliance. This is compared against measured average daily consumption for each customer, drawing on a year's worth of electricity data from customers that had been connected for between 7–15 months. The majority of households (70%) gave an over-estimate of their future electricity use, and there was a very large mean absolute error of 426 Wh/day per customer on an average consumption of 113 Wh/day per customer. By not setting a time frame for the acquisition of aspirational appliances (or not probing for these data from

customers), the expected electricity demand for the mini-grids was vastly over-estimated for the 15 months after installation. In addition, the rate of load growth was just 8% yearly (calculated over the first two years), which does not suggest the rapid load growth required so that actual electricity use better reflects predicted use (though of course load growth in these contexts can be unpredictable, and the trend showed large fluctuation). In general, this study shows that households are not good at estimating what appliances they will have in the future and how they will use them, but also calls into question the usefulness of the energy-use survey methodology as used for these sites to accurately predict demand.

Hartvigsson et al. [86] pursue a similar study for a mini-grid in Tanzania, but the energy-use survey is conducted post-installation and so is indicative of respondents' attempts to estimate their current electricity use, rather than aspirational electricity use. Demand estimation for the whole system, households, and small to medium enterprises are considered separately. In contrast to Blodgett et al. [44], they find that energy-use surveys under-estimate the actual total daily electricity demand, by 48-117% for the three categories. They find the largest discrepancy occurs during the night, suggesting respondents are particularly bad at estimating night-time consumption. However, the results are not as robust as those of Blodgett et al., because having conducted energy-use surveys for 28 households and 19 enterprises, the comparison to measured data is based on electricity data from just one household and one enterprise chosen as typical among those surveyed. It therefore does not give as detailed an account as Blodgett et al., but still shows an accuracy issue with relying on self-reported assessments of behaviour for demand estimation calculations.

2.2.2 Demand Estimation from Measured Electricity Data

There are various approaches to using existing measured electricity data to estimate demand for new un-electrified communities. In this section we cover the direct transfer of data from one project to another, and two methods that relate descriptive data about the end user to the electricity data.

2.2.2.1 Direct Transfer from Existing Projects

The most basic approach is simply the application of data measured from one project or area to another. In practice, this is hampered by the lack of data that exist about electricity use of low-income, rural people. Some research does present electricity data from case

studies of electricity access systems. Often these data sets are not detailed data on household electricity use; instead, they are usually aggregated electricity data at system level, and/or average electricity use [44], [46], [86]–[89], so do not provide detail on the variance and diversity of electricity use in this context. Part of the reason that detailed data sets are not released is that there are privacy concerns with the release of high resolution electricity use data [87]. Another reason could be that these data are valuable to system developers whose systems generate it, so they do not want to release it. The use of existing data occurs more frequently when the data sets concerned are census or large-scale survey data, as discussed above for estimating appliance use in households [38], [82].

Apart from the challenge of a lack of data, transferring data directly between projects or areas involves the assumption that the data are representative of electricity use at the target community, despite being gathered somewhere else, at a different time. This was explored by Blodgett et al. [44] following the analysis that showed that energy-use surveys conducted by the system developer gave a large over-estimate of the actual electricity demand for eight mini-grids. The authors considered a ‘data-driven proxy village approach’, to see if using measured data from one mini-grid to estimate demand for another was more accurate than the energy-use surveys. Using pairwise ANOVA tests for the eight mini-grids, with the null hypothesis for each pair that the actual mean daily electricity consumption per customer were the same, only 3 out of 28 pairs were significant at the 5% level. For the 25 other pairs, the null could not be rejected, and the mean electricity use per customer was statistically similar (one mini-grid was common to the three pairs that were statistically different). Consequently they found their data-driven proxy village approach to be more accurate than the energy-use surveys. Estimating mean daily electricity use for each micro-grid using the mean daily electricity use of the other micro-grids, gave a mean absolute error of 75Wh/day per customer (66% of average daily consumption), across all pairwise combinations. This is in comparison to a mean absolute error of 426 Wh/day per customer (377% of average daily consumption) for the energy-use surveys method. They conclude that developers would get better estimations of electricity demand from real-world measured data sets, and hypothesize that weighting the results based on customer characteristics would improve the results.

2.2.2.2 Econometric Models

Econometric models often use regression analysis to explain electricity consumption with socioeconomic and demographic explanatory variables. The resulting model can then be applied to estimate demand for other households or geographical areas, when data are available on the explanatory variables.

In Louw et al. [90], the drivers of consumption of electricity are investigated for newly electrified low-income rural sites in South Africa. Income, the price of alternative fuels, appliance ownership, and access to credit were statistically significant in the prediction of electricity demand. They report an R^2 (coefficient of determination - the proportion of variation in the dependent variable explained by the model) of 0.4 for the best models tested, showing that the majority of the variance in electricity use is still unexplained by these variables. This is not necessarily surprising, because human behaviour is complex, the number of explanatory variables possible to test is limited by data availability, and some drivers of electricity use may not be possible to measure directly (e.g.: abstract aspects of culture).

In Otieno et al. [91], data from four micro-grids in East Africa are used to regress average daily customer electricity use onto explanatory variables about the customer. Income, type of appliances present in the home, fuel use, mobile phone carrier, and latitude were found to be significant, and similar R^2 values were reported as for Louw et al. [90]. However, a second part of this work uses a similar regression to model the average daily electricity use, but uses explanatory variables extracted from the load profile, such as consumption at a certain hour. The purpose of this is unclear, as it is estimating a variable (average daily electricity use) using data extracted from the same observation in a different form (the load profile). Unsurprisingly, a better model is found ($R^2=0.73$), but the exercise itself does not make sound sense.

A further example of these models in Global South electricity access contexts is Zeyringer et al. [92], where a regression analysis of monthly average demand for areas of Kenya is used to determine whether grid-extension or standalone solar PV systems are the most cost-effective strategy.

The strength of the econometric model approach is that one can investigate the significance of numerous explanatory variables on aggregate measures of electricity use and identify what variables are most important. The downside of the approach is that it provides an estimate of the conditional mean of an electricity metric for each household (for [90], [91]), or each

geographical area (for [92]), rather than indicating the distribution of electricity use or the load profile. Prediction intervals can be calculated, but this assumes normality. A further drawback is that it requires data for numerous explanatory variables to be available, which may be particularly unlikely on the scale of individual household. If the data are not available, gathering it can be costly in time and money.

2.2.2.3 Load Profile Classification Models

Load profile classification models to estimate electricity demand are most frequently seen in Global North contexts, and seek to classify customers according to the load profiles that represent their electricity use patterns. There are various motivations. Utilities may want to characterize users by electricity use so they can determine what tariff to offer customers [93]. Linking the groupings to measurable customer data (data about business type, size, demographic and socio-economic data) can provide insight on what demand side management initiatives would be appropriate for different groups, and how to engage users with more sustainable practices, [93], [94]. Results can strengthen energy modelling and forecasting and provide insight on energy policy matters [95]. These motivations differ from that of this research (to estimate demand for un-electrified communities), because in Global North contexts, almost everyone is already an electricity customer and the focus is management of the customers. However, the principle of linking load profiles to demographic data is useful to examine, and, as will be shown, there are very few examples of this on Global South electricity data.

Typically, the first step is segmentation of users according to their electricity use. Clustering analysis is usually used for this, which is a data analysis technique that involves grouping similar data points together. The aim is to find a grouping that maximizes differences between the groups while minimizing differences within the groups. In the context of time series electricity data, clustering is used to group together load profiles that are similar, splitting the set into a smaller number of representative groups [96]. The data set of load profiles is then summarized by a smaller representative set, which is usually the centroids (mean) of the groups (these are often called ‘representative load profiles’).

If each load profile comes from a different customer, the clustering of load profiles groups the customers according to their load profile. Data that describes the type of customer can then be linked to the groups. The simplest way is an inspection of the groups with respect to

the data of interest. In Gouveia et al. [97], clusters of residential households are inspected in terms of survey data, and a qualitative description of each group given.

In other cases, modelling of the link between groups and survey data is attempted. Often, regression models are used to explore which explanatory variables are significant in the classification of a customer to a group [94], [95], [98]. Other classification models can be based on decision tree algorithms [93]. In each case, the classification model accuracy is evaluated and the explanatory variables that are most significant in the classification are found. The models can then be used to classify new customers. For example, in Vercamer et al. [93], commercial customers are classified according to data about the company, and in Viegas et al. [94], residential customers in Ireland are classified using survey data and short extracts of smart meter data.

Due to the lack of data in the rural Global South context, there is little similar research using a segmentation step and then a classification to estimate the load profile for new connections. Three relevant papers are reviewed here.

In Williams et al. [99], clustering analysis is used on customer electricity data from four micro-grids in East Africa to group average customer load profiles and average daily electricity consumption (in separate exercises). The results are 15 categories of users that are then inspected to see what type of customers are in each category (home, business, both, public premises). In Otieno et al. [91], the same micro-grid electricity data are grouped by clustering on a combined data set, consisting of demographic data and features extracted from load profiles, to investigate the differences in electricity use between households and businesses. This gave three groups of customers, and businesses and households were found to be intermingled through all three clusters. Both papers give useful insight into the different types of electricity use for different categories of customers for that particular data set, and extension into a predictive model for classification of a certain customer type would be interesting.

Llanos et al. [100] do contribute a combined segmentation and classification step for rural, low-income customers in Global South contexts. The purpose is to estimate electricity demand for an un-electrified community, having first analyzed data from an electrified community. They cluster household load profiles from an electrified community into groups, and characterize the groups according to survey data about the characteristics of the households within them. Example groups are ‘elderly couple, adult couple living with two young children’. These household types and associated representative load profiles

form a database. To apply these profiles to a new un-electrified community, demographic data are gathered for the target community. Another clustering process is used on the survey data to group households together, giving the ‘types’ of households in the new community. These household types from the un-electrified target community are matched to the types in the database, using a ‘heuristic selection technique’ which seeks the closest match between types. The associated representative load profile is then assigned to the group from the new community. The overall system load profile is estimated for the project by aggregating the household load profiles together. This method relies on it being possible to describe the groups of households, clustered by their load profiles, with one description summarizing their demographic characteristics, and this in itself requires (what could be) subjective user input. It is also relatively labour intensive, requiring collection of multiple variables of demographic data, two clustering processes on different data (only two clustering algorithms are investigated), and a somewhat vague ‘heuristic selection technique’. However, it is a useful contribution in the Global South context to estimate electricity demand through the transfer data from an electrified community, to an un-electrified community, through observations of social and demographic data.

However, the drawback to these load profile classification models reviewed above for the Global North, and for Llanos et al. [100], is that they associate customers with just one load profile. The issue with assigning one load profile for each customer is that it does not take account of the variation in electricity use that exists across the days of each customer. This is important to consider in the design process because designing systems for the average use may mean that higher demand scenarios can not be met. These scenarios may nevertheless be important use patterns for the end user. Also, when households are aggregated together, knowing variance in household use enables the effect of demand diversity to be taken into account [40].

Association with one load profile is a result of the strategy of the clustering analysis used to achieve the segmentation of users by grouping similar load profiles together. The issue of what strategy to take in a clustering analysis is discussed in the next sub-section because it is relevant to the directions and decisions underpinning this research in the demand estimation topic. In Section 2.3.1, further theory behind the application of clustering algorithms is covered.

2.2.2.4 Influence of Clustering Analysis Strategy

To perform a clustering analysis on electricity data from multiple customers in order to separate the customers according to their electricity behaviour, the form of the electricity data contributed by each customer must be determined. These are data from customers that are to be compared in the clustering analysis. The raw data for each customer most often consists of a set of metered electricity data, recorded at defined intervals throughout the day, for multiple days. The most common approach to define the features for clustering is to calculate a customer representative load profile (RLP) by averaging across a selection of the days measured for that customer for each time interval in the day [96]. This gives an average load profile, spanning one day, for each customer. The single customer RLPs are then grouped using a clustering algorithm, so that each customer is assigned to one cluster.

Some clustering strategies seek to take account of the variation within users by allowing multiple RLPs to represent each customer. This can be achieved through a two-stage clustering process such as in [101], [102]. The first stage is a clustering of the days for each customer individually, to find a number of customer-specific RLPs. This shows the typical patterns of electricity use that describe the electricity use for each customer. For the second clustering stage, one, or all, of the customer-specific RLPs can be compared across all customers and clustered into groups. In Tsekouras et al. [101], the RLP that represents the most common customer use pattern is carried forward to the second stage of clustering, where it is compared against the most popular RLPs from the other customers. However, this still means that each customer is represented by just one cluster after the second clustering process, and the variation in use is not captured. In Mets et al. [102], all customer RLPs are carried forward into the second clustering stage, compared to each other, and grouped into a set of ‘global’ clusters. Each customer is represented by numerous global clusters, given by a vector which describes the number of days of each customer represented by each global cluster. In this way, the variation in individual customer electricity use is retained throughout the clustering process, and customers can be compared according to the combination of global clusters they belong to, and the proportion of their usage patterns represented by each. This two-stage strategy is far less common than assigning a single load profile to each customer, and is a more complex process and computation. However, the advantages of retaining the variation in use for each customer is much more detailed information about the electricity use patterns in the data set.

2.2.3 Summary

This review has shown that there is a need for accurate demand estimation methods for un-electrified rural customers, to replace the standard industry practice of energy-use surveys. Two case studies investigating the accuracy of energy-use surveys found them to be inaccurate. However, the direction of inaccuracy differed between the two: in the study by Blodgett et al. [44], an over-estimation of electricity demand is found, and in Hartvigsson et al. [86], an under-estimation is reported.

To provide improved models to estimate electricity use in this context, there are methods that link electricity use to descriptive data about the customers. Econometric models typically regress a single metric of electricity use using explanatory variables about the customer, looking for determinants of the magnitude of electricity use. Load profile classification models seek to characterize customer according to their typical pattern of electricity use, and in some cases find a prediction model for the classification from customer data. The majority of work in this area has occurred for Global North contexts. Where Global South data have been analyzed, only rarely has a predictive model been found such that load profiles for un-electrified communities can be estimated (as far as the author knows, Llanos et al. [100] is the only evidence of this).

A downside to both the econometric models and the load classification models is that they associate each customer with just one measure of electricity demand, be it an aggregate measure or load profile, and do not take account of the variation in electricity use for each customer. For load profiles, no research has investigated two-stage clustering analysis approaches for Global South electricity data, which allow the variation in each customer's electricity use to be retained through the analysis. A further drawback is that the regression methods are most effective when it is possible to test a number of potential explanatory variables for significance in the regression or classification, but it may be costly to gather these data. In particular, data on income, which is found to be a significant determinant of electricity use in Louw et al. [90], can be difficult to obtain as respondents are unwilling to disclose financial information [81].

2.3 Part 3: Background Theory

2.3.1 Clustering Analysis Applied to Electricity Data

This section provides background theory to clustering analysis of electricity data, covering the input data format and preparation, the various algorithms, and how to choose a particular algorithm and solution. The strategy taken to apply clustering analysis was reviewed in Section 2.2.2.4 because it is relevant to the review of demand estimation methods.

Clustering analysis is an unsupervised classification method, because the groups found in the data are not known *a priori* - they are derived from the data. There are numerous clustering algorithms, and they have different properties that define similarity differently and result in different groupings of input data. Clustering analysis is commonly used in the analysis of electricity use patterns to group together electricity data that show similar electricity behaviour. It is used in this research to find typical load profiles of electricity use within and across households, and this review is relevant to the analysis in Chapter 6.

2.3.1.1 Input Data

Clustering analysis can be performed on time domain electricity data, or on features that are derived from the time domain data [96]. If it is performed on time domain electricity data, each ‘data point’ to be clustered is usually a load profile, defined by a number of power or energy values that describe the electricity use over a set time period, most commonly a day. For example, a data point would consist of 24 values of energy, one for each hour of a day. If features are extracted from the time domain data, they typically describe some aspect of the shape of the profile such as mean or peak energy use [103]. An argument for feature-based clustering is that it reduces the dimensions of the input data in comparison to time series data. However, there are methods to reduce dimensions for time series data (see the next section), and in Williams [104], time domain data gave better results than feature-based models according to a number of evaluation indices.

Some approaches to the analysis perform clustering on both load profiles and features. This occurs in Williams et al. [99], where clustering is carried out on the normalized load profiles and the average daily electricity consumption separately. This resulted in three clusters of average daily electricity consumption, and five clusters of load profiles, and thus 15

customer groupings. This is a high number of groups to meaningfully visualize and interpret, which is often part of the objective of carrying out a clustering analysis.

2.3.1.2 Pre-clustering Steps

Normalization of the input data and reducing the number of dimensions before clustering are standard pre-clustering steps. The information in the following section can be found more fully in [105].

Data normalization is required when variables to be clustered have different measurement scales and ranges, such as height and weight of humans. In addition, Euclidean distance can accentuate some variables over others if there are large differences in the magnitudes of the variables. For electricity data, normalization can be used either to avoid this, or purposefully to compare shape rather than magnitude of the load profiles. A common approach for electricity data is to normalize load profiles so that the minimum and maximum power values are 0 and 1 [96]. If the relative magnitude across the load profiles is of interest, and if there is low variation in magnitude of electricity used, normalization should not be used.

Principal component analysis (PCA) is a common way to reduce the number of dimensions of the input data set to simplify a clustering computation while preserving the variance in the data. When using PCA for dimension reduction, the aim is to project input data which consists of p -dimensional feature vectors, to a q -dimensional subspace defined by q principle components, where $q < p$.

Let the input data be a matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, consisting of n observations and p features, and where each feature has a mean of 0. PCA involves computing the eigenvectors of the covariance matrix of $\mathbf{X}$, which are the principle components:

$$\mathbf{V} = \text{eigvec}(\text{cov}(\mathbf{X})) \quad (2.1)$$

where: $\mathbf{V} \in \mathbb{R}^{p \times p}$, with p principle components/eigenvectors as columns, ordered in decreasing size of the corresponding eigenvalues; $\text{cov}()$ is the covariance function. The first principle component (the first column vector in $\mathbf{V}$) is the direction in the data with the most variance, the second is orthogonal to this and is the direction in the data with the next largest variance, and so on.

A subset, q , of the principle components are selected, which are the q ‘largest’ principle components in terms of the variance of the data they represent. The value of q is chosen

according to the percentage of variance that one requires to be retained in the transformation.

$\mathbf{V}_q \in \mathbb{R}^{p \times q}$, and consists of these principle components.

The original input data are then transformed:

$$\tilde{\mathbf{X}} = \mathbf{X}\mathbf{V}_q \quad (2.2)$$

where $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times q}$, consisting of the n observations transformed into the subspace defined by the q principle components.

Clustering analysis is then performed on $\tilde{\mathbf{X}}$, and the results are visualized in the original feature space for evaluation.

2.3.1.3 Clustering Methods

Table 2.2 compares six clustering algorithms that have been used for clustering electricity data. The first reference after the name of the algorithm is the primary reference that describes the algorithm. The references following it are literature where the algorithm has been used in the clustering of electricity data. Although self-organizing (SOM) are primarily known as a tool for dimension reduction, they are a popular choice for clustering electricity data so are included.

For hierarchical agglomerative clustering (HAC), a key parameter that must be chosen is the linkage criterion. Table 2.3 compares four of the most common criteria.

For three of the algorithms (HAC, DBSCAN, spectral clustering), the distance metric must be chosen, defining how the distance between two points is measured. The choice of distance metric can change the shape of the clusters found, and some studies investigate a number of distance metrics for comparison, such as in Williams [104]. Williams found that the Euclidean distance produced slightly better solutions than other distances tested in that work according to several internal clustering quality scores (discussed in the following section). This work also uses the Euclidean distance, which is one of the most common distance metrics, defined for points a and b , which are vectors of length p , as:

$$d(a, b) = \left(\sum_{i=1}^p |a_i - b_i|^2 \right)^{\frac{1}{2}} \quad (2.3)$$

Different algorithms seek clusters in different ways: by minimizing total within cluster variance (k-means), according to the similarity of two clusters or data points (HAC, affinity

Table 2.2: Review of clustering algorithms: Algorithm, associated parameters, variations, and strengths and weaknesses

Name	How it works	Parameters to choose	Variations	Strengths/ Weaknesses
K-means [105], [95], [96], [98], [99], [101], [106]–[112]	For a set of observations x, find k sets $S = S_1, S_2, \dots, S_k$ that minimize the inertia: $\arg \min_S \sum_{i=0, x \in S_i}^k x - \mu_i \quad (2.4)$ where μ_i is the mean of S_i, also known as the centroid of cluster S_i. Start with initial random centroid placement, and assign each point to closest centroid. Recalculate centroids as the mean of all assigned points. Re-assign points to centroids, re-calculate centroids, until stable.	k: Number of clusters	K-medoids: centroids are data points, uses the Manhattan distance [98], [107]. Fuzzy k-means: each data point has a membership degree to each centroid, and ‘fuzziness’ is controlled by a parameter. [96], [113], [114] k-means++: Improves initialization of centroids by initializing them far away from each other [115]. Adaptive K-means: A mean squared error threshold for a cluster means new centroids are generated when it is violated (choose threshold rather than k) [116].	Poor with clusters that are non-convex/ anisotropic/ not clearly defined. Have to choose k. Sensitive to centroid initialization - though k-means++ improves this.
Hierarchical Agglomerative Clustering (HAC) [105], [96], [97], [101], [107], [110], [116]	Starting with each point in its own cluster, at each step the two most similar clusters are merged, until all data points are in one cluster.	Linkage criteria: defines the similarity between two clusters (Table 2.3). Distance metric: defines the way distance between two points is measured. k: number of clusters.	Divisive: starting with one cluster containing all data points, clusters are progressively split. Each time, the cluster with the largest internal dissimilarity is split.	Can handle anisotropic and non-convex clusters. Different linkage criteria can be used to influence cluster shapes. Makes a locally optimal decision at each step, which does not necessarily result in a global optimum for any k.
DBSCAN (Density-based spatial clustering of applications with noise) [105], [107], [109]	Finds clusters of high density, separating them from low-density areas. Each point is tested to see how many data points lie within ϵ distance. If there are x data points this becomes a cluster.	ϵ: distance threshold, to determine what points are close enough to form a cluster. Distance metric: See HAC. x: minimum number of samples that constitute a cluster.	N/A	Can cope with clusters that are not convex, but not good with clusters of different densities. Do not need to choose k - but modification of ϵ and x will influence number of clusters

Continued on next page

Continuation of Table 2.2

Method	How it works	Parameters to choose	Variations	Strengths/ Weaknesses
Spectral Clustering [117], [93], [118]	Family of methods that use the eigenvectors of the similarity matrix in the clustering process.	Distance metric: See HAC. k: number of clusters	There is variation in which eigenvectors to use and how to derive clusters from them, see [119] for a review.	Clusters by affinity (connectivity), so good at non-convex clusters.
Affinity Propagation [120], [121], [122]	Finds 'exemplars' which are data points that are representative of clusters. 'Messages' are sent between pairs of points, updating the suitability of each point to be an exemplar for other points according to what's happening with surrounding points. Repeats until convergence.	Preference: set globally or customized for data points that should be exemplars. Higher global preference means more clusters. Damping factor: damps messages between data points to avoid oscillation. Similarity: similarity matrix required to start the algorithm.	N/A	Do not need to choose k - but the preference must be chosen, influencing the number of clusters.
Self-organizing maps (SOM) [123], [98], [100], [110], [111], [124]	A type of unsupervised neural network. Comprises nodes/ neurons in a (usually) 2-dimensional grid that adapt to the shape of the input data, so a higher density of nodes exist where there are more data, showing clusters. The grid is randomly initialized in the data space. Picking each input data point in turn, the nearest node is the 'winner'. It moves closer to the data point, and its neighbours move closer by lesser degrees. The amount moved by the nodes, and the neighbours affected, decreases after each iteration. Continue until stable.	N: Number of nodes, e.g.: choose N_1 and N_2 for $N = N_1 \times N_2$ nodes. Learning rate: Distance to move winner node towards the input data point. Neighbourhood function/ radius: Determines how neighbouring nodes are affected and size of area around each winner node affected. Reduction rate: Rate of decrease in learning rate and neighbourhood radius.	Different ways of using a SOM: in [124], the N is larger than the number of input samples, and clusters found from the resulting SOM by visual inspection/ k-means. In [98], N is set so number of nodes is equal to the number of clusters sought.	Not good at categorical variables, and the initial position of neurons is randomized each time, so results can be inconsistent. Useful for reduction of dimensions, and when used in 2D effective for data visualization.

Table 2.3: Linkage criteria for hierarchical agglomerative clustering, demonstrated for two sets of observations A and B . Reference: [105]

Linkage Criteria:	Single	Complete	Average (UPGMA)	Ward's
Diagram				
Proximity definition	$\min\{d(a, b) : a \in A, b \in B\} \quad (2.5)$ where $d(a, b)$ is the distance between points a and b. 'Nearest neighbour' – distance between two closest data points.	$\max\{d(a, b) : a \in A, b \in B\} \quad (2.6)$ 'Furthest neighbour' – distance between two furthest data points.	$\frac{1}{ A B } \sum_{a \in A} \sum_{b \in B} d(a, b) \quad (2.7)$ 'Unweighted Pair Group Method with Arithmetic Mean'. Takes an average of all the pairwise distance.	$\min \Delta \left\{ \sum_{i=0}^k \sum_{x \in S_i} x - \mu_i \right\} \quad (2.8)$ where, for a clustering of observations $x = (x_1, x_2, \dots, x_n)$, k is the number of clusters S_i, and μ_i is the mean of S_i. This criterion takes the agglomeration step that minimizes the increase in the within cluster sum of squared residuals.
Typical behaviour	Prone to a 'chaining' effect, whereby data points are attached one by one to clusters. This means it is less likely to produce reasonably sized groups at later stages of the algorithm but might do better finding elongated or anisotropic clusters.	Forces more spherical clusters than single linkage, which have roughly equal cluster diameter.	A compromise between single and complete - clusters with a range of sizes.	Shares the same objective function as k-means, but may find a less optimal solution as it cannot change previous assignments. Tends to find clusters of reasonably equal sizes, like k-means.

propagation), by distinguishing data according to density (DBSCAN), by mapping a neural network onto input data (SOM). This means they find different solutions in the same input data. A key point of comparison between methods is what shapes of clusters an algorithm is able to detect - for example, k-means works best with spherical clusters of roughly equal size, which are isotropic, and is not good at detecting non-convex clusters. Other methods, like HAC, DBSCAN, and spectral clustering, can detect non-convex clusters because connectivity of data points is important in those algorithms. Within HAC, the different linkage criteria can be used to find a range of clustering outputs, from isolation of the most different days (single), to more equally sized clusters (complete) (see Table 2.3).

Most often, a single clustering algorithm is used to cluster the data. More rarely, methods are used in sequence, for example as in [116], where a variation of the k-means algorithm is applied first, and then HAC used to merge cluster centroids that are close together.

2.3.1.4 Choosing a Clustering Solution

Determining a clustering solution requires choosing the algorithm and any associated parameters, which often includes choosing the number of clusters. The desire to find the ‘best’ or the ‘right’ clustering of a particular data set is very common.

An important consideration that can be used to guide the choice of the algorithm is how the input data are arranged in the input space. Some algorithms are weak at recognizing certain shapes, and so if the way the data are arranged can be visualized, a choice can be made according to the algorithm properties. However, for data with more than four dimensions, visualization can be difficult, although data reduction techniques such as PCA can be used to project the data onto a smaller number of dimensions.

An alternative, or additional, means to choose a clustering solution, is to use internal clustering quality scores. These scores seek to quantify the clustering solution, and are based on aspects of the clustering shape, such as the dispersion, compactness, or distinctness of clusters. They are ‘internal’ because they depend only on the data and the shape of the clustering solution. Table 2 in [96] gives an overview of 10 such scores. They are used frequently in the literature to choose a method or a particular solution within a method. However, it is commonly noted that different scores give different answers as to the ‘best’ number of clusters and cluster method [104]. This is because they emphasize different aspects of the clustering solution shape as ‘good’, depending on the definition of the score. In

addition, scores will be biased towards clustering algorithms if they measure a quantity that is correlated with the objective function of the algorithm, or is in another way similar to it. For example, judging a clustering result using the inertia will show k-means to be better than hierarchical methods, because k-means seeks the clustering solution that minimizes inertia. Because different scores give different answers, it is common practice to use a selection of them to evaluate a clustering solution [104], [107], [125].

Another alternative approach to the choice of a ‘best’ algorithm or particular solution is discussed in the insightful paper, ‘Clustering: Science or Art?’, by von Luxburg et al. [126]. They argue that, in fact, clustering is not an application independent mathematical problem, and that domain independent means (such as internal quality scores) cannot indicate which algorithm is the right one. They write that ‘depending on the use to which a clustering is to be put, the same clustering can either be helpful or useless’. The ‘best’ clustering solution depends on why the clustering is being carried out, and what will be done with the output. They argue that visualization of the results is important, so that the solution can be judged in the context of the problem. Internal quality scores are more useful when applied to solutions within a certain algorithm, if the score measures the aspect of clustering that is important in the application.

The first and third approaches described above work well together to address the ‘best solution’ issue. The school of thought that a clustering solution should be judged in the context of the problem being addressed is particularly important when there is not a visible ‘natural’ segmentation of the input data. A lack of a clear grouping is likely for real-world data, particularly data produced by human behaviour, like electricity data. This is because human electricity behaviour is likely to occur in a spectrum, rather than fall into very defined groups. Assessment of this is possible by visualizing the input data, which was the recommendation in the first approach. If, for example, there are four discernible groups in the input space, this may indicate a ground-truth grouping of the data. It may not even be necessary to use a clustering algorithm to segment the data if the groups are very clearly defined. If an algorithm is required, though, the choice can be made according to the shapes in the input data - guiding whether an algorithm that reduces variance is required (e.g.: k-means), or if one that emphasizes connectivity of data points is more suitable (e.g.: spectral clustering). However, if there is no clear grouping, or if it is not possible to visualize the data, the third approach discussed here becomes important. There may not be a natural grouping of the data,

and therefore the best clustering solution should be judged against criteria that are relevant for the context and why the analysis is being carried out.

In the literature, a mixture of approaches is common. Some authors use internal quality scores, and also evaluate the clustering by visualizing the results and considering them in perspective of the problem they aim to address. For example, Viegas et al. [94] uses three scores to evaluate k-means clustering results of load profiles for different k values, resulting in three different answers for optimum k value. For each k , the cluster centroids and associated profiles are plotted, and looked at by eye to identify the solution that has compact clusters, but also a relatively even number of profiles in each cluster. Other authors choose the solution according to their problem definition and context without using any internal quality scores. In Yamaguchi et al. [127], the k values to investigate using k-means clustering are first limited to below $k = 20$, because more customer groups are not useful to electricity suppliers. After that, the final k solution is selected by weighing up that fewer clusters are advantageous for the regression model the analysis will be used for, but more clusters are helpful to capture variation in the data.

2.3.2 The Energy Cultures Framework

The Energy Cultures Framework (ECF) is a tool to understand energy behaviour and its determining factors. It focuses on three key aspects: 1) Material culture (physical objects/infrastructure), 2) Practices (actions by the actors concerned) and 3) Norms (expectations, which are reflected in current practices and material culture, and aspirations, which have not yet been achieved). It is multi-scalar, such that it can be applied to an individual, a community, a business, or a sector. A particular ‘energy culture’ is characterized by a subject’s norms, practices, and material culture, interactions between them, and includes the influence of external factors outside of the subject’s agency or immediate surroundings, as represented in Figure 2.2.

The ECF was developed by a multi-disciplinary team from the University of Otago, New Zealand as a means to understand behavioural change with a multi-disciplinary approach [47]. It was first developed as part of a research project to understand why households in New Zealand were slow to adopt energy efficiency measures [128]. It has since been used to analyze processes of cultural change, such as the increased use of solar lighting in Vanuatu [129], and changing mobility habits of young people [130]. It has also been used to develop

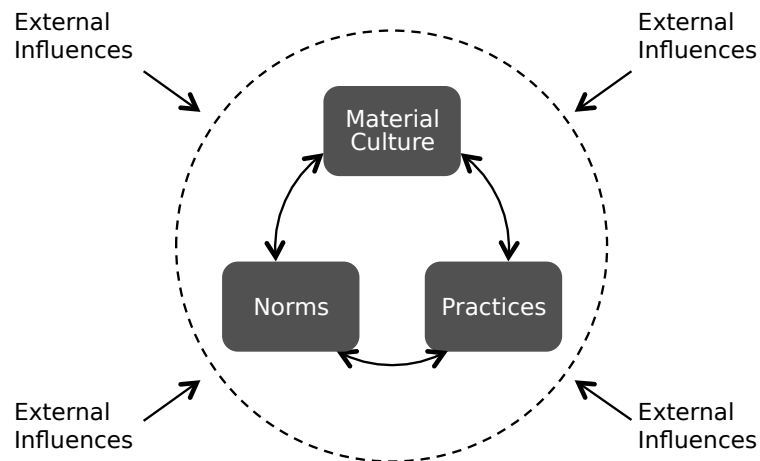


Figure 2.2: Schematic of the Energy Cultures Framework. From [47].

recommendations for policy interventions to encourage change of behaviours in residential energy use [131]. The advantage of the framework is that it assists multi-disciplinary teams to work on a common problem or project, providing a common language and integrating model of behaviour and context that is accessible to all disciplines.

In this research, the ECF is used to support the interpretation of different patterns of electricity use in the households in the case study. For this purpose, it is useful to introduce the factors that contribute to energy behaviour in rural, low-income, Global South contexts, in terms of the material culture, practices, and norms, of the households in the case study. Figure 2.3 shows the factors that the solar nano-grid research team deemed important for this context, grouped by the aspect of the ECF they are associated with. Community members' material culture covers the physical matters such as the fuels used, the appliances they have, and the characteristics of their house. Their practices comprise their behavioural characteristics relating to energy use, such as how they obtain fuel, what they pay for the fuel, their practical know-how, and how they use appliances. A person's norms encompass their expectations about how they want to live their everyday life, and their aspirations for change including desire for particular appliances or for higher social standing or influence in their community. In addition norms include their ambitions in the broadest sense, their relevant beliefs and values, and any health and safety concerns towards different energy sources.

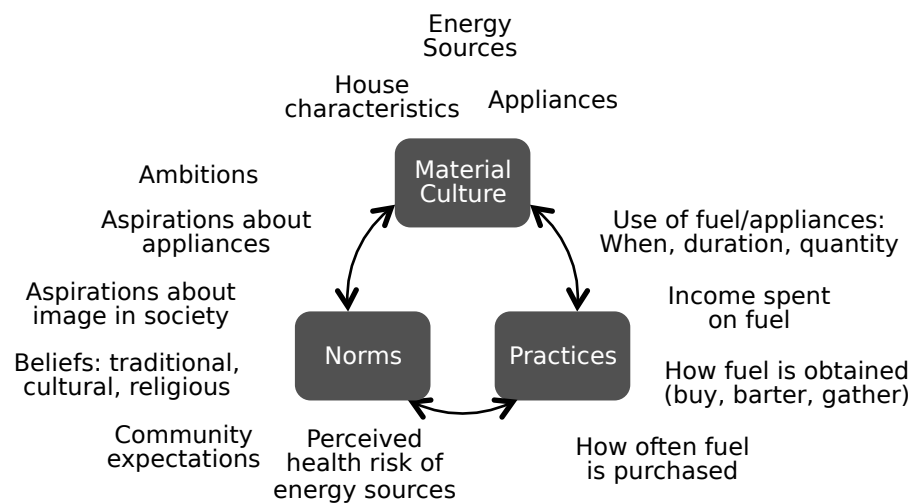


Figure 2.3: The Energy Cultures Framework applied to a rural village in the Global South. Factors of material culture, practices, and norms.

To build electricity access systems that are suitable for specific contexts, how can we take account of end-user needs, aspirations, and context, in the design process?

3

The Service Value Method for Design of Electricity Access Systems

Contents

3.1	Introduction	54
3.2	Service Value Method	56
3.2.1	Step 1: Data Gathering: Service Value Test (SVT)	57
3.2.2	Step 2: Data Interpretation: Service Maps	60
3.2.3	Step 3: Preliminary System Design	64
3.3	Case Study: Rural Electrification System Design in Kenya and Bangladesh	65
3.3.1	Step 1: Data Gathering: SVT Implementation	65
3.3.2	Step 2: Data Interpretation: Results	65
3.3.3	Step 3: Preliminary Design: Solar Nano-grid	68
3.3.4	Solar Nano-grid Implementation	72
3.4	Discussion and Evaluation	74
3.5	Summary	78

3.1 Introduction

To improve the usefulness and successful implementation of electricity access systems for the rural Global South, end-user needs, aspirations and wider contextual factors must be identified and included in the design process. Design teams are frequently working outside of contexts familiar to them for electricity access projects, and methods must be accessible for disciplines that are not trained in social sciences. This chapter presents the Service Value Method (SVM) as a means to gather the required data and integrate it into the design

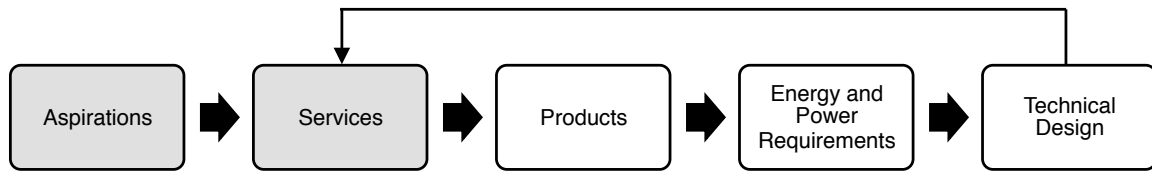


Figure 3.1: Linear representation of design phases. Any phases could be returned to in an iterative design process in practice, with an example feedback loop from technical design to services shown here. Grayed boxes represent the Service Value Method scope.

early on in the design process.

The appreciation of the importance of including end-user needs and context must translate into the design approach for electricity-enabled services. The SVM is based on understanding the design phases for electricity access systems as represented in Figure 3.1. There is an absence of known demand, established infrastructure, and standards and specifications for the off-grid Global South context. Therefore, we propose that designers of electrical systems should start by understanding end-user aspirations and the services that meet them, alongside consideration of contextual factors, to incorporate these aspects throughout the design process. Using a service-oriented approach complementary to the sustainable product-service systems concept, the SVM focuses on the first two shaded phases of design in Figure 3.1. End-user needs, aspirations and context are uncovered through understanding important services from the perspectives of end users, and these are integrated into the design process.

Data are gathered through a rapid but effective field exercise implemented through focus groups with end users, using a three step qualitative-quantitative-qualitative approach to probe for end-user aspirations and context. Once the data are gathered, there is a structured method to interpret and integrate the data into the design process, which enables the preliminary design of the system. The scope of the SVM is to gather data sufficient to support preliminary design level decisions by determining what services should be provided by the electricity system. In addition, to maintain the method as efficient and easy to replicate, the focus is on immediate end-user context, such as access to markets, practices and material flows in and around the village. This means it is not intended to provide all the contextual information required for a successful detailed design and implementation of an electricity system. Rather it is a useful and efficient scoping exercise that ensures end-user needs and context are included early on in the design process. It should be positioned within a broader design process, which might include

iterative design cycles, and various methods of participation, and in conjunction with other methods to address contextual factors flagged by the SVM and broader context in more detail.

Section 3.2 describes the steps of the SVM and the underlying principles. The SVM was used to design solar nano-grids for two villages in Kenya, and the data-gathering process was used in two villages in Bangladesh to contribute to a similar design process there. The results of this case study are presented in 3.3. Finally, Section 3.4 discusses the strengths and weaknesses of the method, drawing on the case study experiences.

3.2 Service Value Method

The Service Value Method has the following aims:

- To gather data on end-user needs, aspirations and surrounding context at a level appropriate for preliminary design
- To enable designers to integrate the data into the design process effectively
- To be fast and effective in field
- To be appropriate for use in Global South contexts

The method does not limit services to be only those enabled by energy or electricity, despite the purpose of the method being to support design of electricity access systems. All services that are desirable to end users are of interest because electricity can underpin and enable a variety of services in ways end users may not be fully aware. Furthermore, services that consist only of infrastructure may provide useful understanding of contextual factors. It is important to understand how energy and electricity services fit into the broader lives of the end users so that a design does not conflict with existing practices they may have.

A distinction is made between ‘household’ services and ‘community’ services to distinguish between services accessed at the homes of end users compared to services accessed more centrally, and open to the whole community. This is useful from the perspective of an electricity system design because the location of the point of delivery of the service matters for design of the transmission and distribution components of the system.

To capture end-user needs and context, the first step in the SVM is a data-gathering step, in the form of focus groups run with end users using an exercise called the Service Value Test. To enable designers to take the data into account in a way that contributes to

a successful system design, the second step is a data interpretation process. Quantitative service value data and qualitative service context data are used to construct service maps, displaying service popularity and evaluating services against key decision criteria for the design. Finally, in the third step the service maps are used to make the key design decisions about what services the system will be designed to support.

3.2.1 Step 1: Data Gathering: Service Value Test (SVT)

Real data must be gathered from end users to enable understanding of their needs and aspirations and an overview of relevant design context. Figure 3.2 outlines the steps that comprise the SVT method which gathers data through focus groups with end users. The following section describes the underlying principles of the methodology, explaining why it is structured as it is. The SVT was developed in collaboration with academic engineers and social scientists, and in-country project partners who were familiar with the local context, to ensure it was appropriate for use in the target rural villages.

3.2.1.1 Principles of the SVT methodology

The methodology of the SVT captures what services end users want to see in their households and communities in the future, to show the designer what services, and therefore what products, are important to consider in the design. The first step of the method asks participants to suggest aspirational services. To avoid influencing participant choices, facilitators should not make their own suggestions (beyond an initial example) as they may not reflect participants true needs and aspirations. To address potential literacy and language barriers, services are captured in pictorial form, and a central service chart is used to collect the suggested services so that participants can keep track of their suggestions.

The SVT method is devised to pick out individual and group preferences in terms of the relative importance of the services, producing quantitative data from which the most important services to the end users can be discerned. The method to do this needs to be quick and efficient to be possible in one focus group gathering, to address the challenges associated with accessing participants in remote or rural areas. It must also be easily understandable to cope with potential low-literacy rates. Therefore, participants use counters to demonstrate the value of each service on the service chart, placing more counters on more important services. Each participant uses counters of a different color so it is possible for facilitators

Focus Group Set-up: 10 participants per focus group, two facilitators, translator if required.

Equipment: A1 Service chart, pictorial service cards (pictures or symbols of anticipated services, some blank), non-permanent marker pen, Blu Tack, A5 service chart index cards, pencils, colored counters, audio recording equipment, camera.

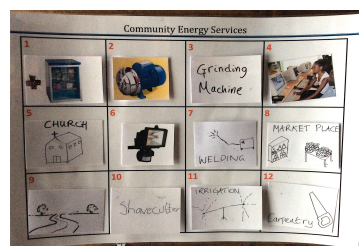
Step 1: Participatory Listing (Qualitative)

Separate SVTs are run for household services and community services.

- Participants are asked to suggest services that they need or want to have in their homes/communities.
- A commonly found example may be used as a demonstration.
- Representative pictorial service cards representing suggestions are attached to the A1 service chart.

Household services are things you want in your home. E.g.: lighting and mobile phone charging.

Community services are things that would benefit the community, that many people might use. E.g.: a health center and a primary school.

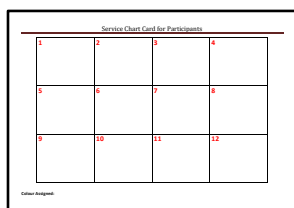


Facilitation questions/statements:

What services would you like to have in your household/community in the next few years?

Step 2: Value Judgment (Quantitative)

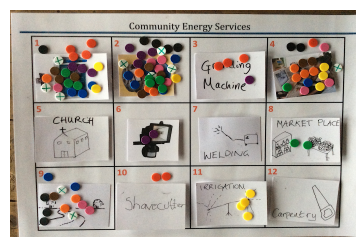
- Individual:** Without conferring, participants indicate service preferences on a service chart index card with crosses using either Free Assignment or Ranking methods.



Options for value quantification:

Free Assignment: Participants place 20 crosses anywhere on the chart with more crosses showing greater importance.

Ranking: Participants allocate 4, 3, 2 & 1 crosses to indicate their 4 most important services.



- Collective:** Participants transfer individual choices to the service chart using colored counters (one color per participant.)

Step 3: Value Exploration (Qualitative)

In a group discussion, explore:

- Reasons for service suggestion and importance, revealing underlying needs and priorities.
- How they envisage using the service (including when), to understand how it might fit into their lives and to reveal associated context.
- How they access the service now (including when), to understand current practices and to reveal more associated context.

How would you use [service]?

When would you use [service]?

How do you access the [service] at the moment?

What else needs to be in place for you to benefit from [service]?

Facilitation questions:

Do you agree with the group results? Why?

Why is [service] important to you/the community?



Figure 3.2: Service Value Test method.

to discern individual preferences and priorities from the whole group response; it also aids participants in keeping track of their own response throughout the exercise. However to avoid

group-think or strong personalities influencing decisions of individuals, participants first make their choices in isolation from the rest of the group, on individual service chart index cards without conferring with each other, before transferring their results to the main service chart.

The emphasis of relative service importance achieved by the distribution of counters on services depends on the method participants use to indicate the value of the services. Free Assignment gives a more balanced indication of a participant's preferences, and group results show less disparity between popular services and less popular services in terms of the number of counters assigned. However, this method is more time intensive and may be difficult for participants with low numerical literacy because of the number of counters/crosses involved. Alternatively, group results from the Ranking method are more likely to result in a few very popular services with the bulk of the counters, and some services with no or very few counters. This can be useful to clearly highlight what the highest priority services are, however there is likely to be less distinction between the less popular services compared to when Free Assignment is used.

The SVT method seeks to reveal underlying needs that motivate the suggestion and valuing of a service through a discussion that enriches the quantitative data from the first two steps. Although experts in their own context, end users may lack an understanding of technology and subsequent opportunities afforded by electricity access. By understanding underlying needs, designers can consider how best to meet those needs through electricity access solutions, services and products that may not be within end users' awareness, and which they therefore are not able to suggest.

The discussion also aims to gather an overview of the context in which services exist, uncovering how end users currently meet the underlying need, how they envisage using the service, what other services are involved, and what barriers might exist to successful utilization. The design must be aligned with current and aspirational practices, and if new practices are introduced, these must not be inconvenient or contrary to current practices to the extent that end users do not engage with the system. Additionally, knowing when in the day the service is required is important for electricity access initiatives using solar generation due to the diurnal irradiance cycle. Energy storage will be needed if electricity is required when generation is not possible. The discussion should be managed to reveal these contextual factors that influence the practicalities of implementing the service and whether or not it will be successful. For example: egg incubators are important for rearing chicks to sell at the

market, but equally important to the success of the venture is access to vaccinations so that chicks survive post-hatching, and the means to get them to a market place.

3.2.1.2 Further SVT Implementation Details

In step 1, if participants contribute products rather than the service that that product enables, it is advisable to clarify what service that product is associated with. Some products can be used in different ways, and end-user intention may differ from facilitators' assumptions. For example, a fridge might be suggested to enable the sale of cool drinks, to store vaccines in a surgery or veterinary setting, or to store food in a domestic or business context. If more than one use of a certain product is suggested, it is up to the facilitators whether these services are represented separately, or together through the suggested product, in which case the discussion in step 3 would seek to understand more about the different uses and their context.

It is advisable to run several focus groups for each project context, separating participants according to demographic traits, to obtain a cross-section of views and ensure fair representation. Due care must be taken that suitable participants are chosen, and that the process is fair and equitable. Care must also be taken to prepare participants and manage their expectations prior to the focus groups and at the start of the sessions before the SVTs are run. In organizing the sessions, the recommended approach is that participants are fully informed of the study and the reasons behind it (for example, that it is part of a design process for an electricity system), but it is strongly emphasized to them that in the focus groups their general opinions and practices are of interest, rather than just those about energy. This is also why the question used in step 1 of the SVT is structured so as to be general rather than targeted probing for important energy services. It is not a problem if participants reflect on their experiences and opinions prior to the focus groups, as long as they have clearly understood the aims of the sessions and do not become biased towards energy-related services.

Further implementation details regarding equipment, participants and process can be found in Appendix A.1 and Appendix A.2, including suggestions to reduce as much as possible barriers to participation due to cultural norms.

3.2.2 Step 2: Data Interpretation: Service Maps

The second stage of the SVM uses the popularity data and service context data from the SVTs to construct service maps. Service maps portray each service, including service popularity,

against key decision criteria for the electricity access project. This section outlines the steps required to process the SVT data and construct the service maps.

3.2.2.1 Identification of Out of Scope Services

Given the SVM uses a broad definition of services, some suggestions from participants may be out of scope for an electricity access project. To determine out of scope services, services are separated into three categories according to the extent to which electricity is a primary component in the service:

1. Services for which no electricity is needed
2. Services which require additional infrastructure alongside an electrical component
3. Services which only require electricity

All services that have no electricity component are judged as out of scope of electricity access projects. However, out of scope services still offer useful information about participants' needs and aspirations, and may enhance understanding of the context of services to which electricity can contribute. Therefore, they are still included on the service maps without association with the decision criteria, with the popularity data displayed.

3.2.2.2 Decision Criteria for Electrical Services

Service maps show each service relative to how well it aligns to key decision criteria for the design of the electricity system. Decision criteria should be chosen according to the project context, guided by the aims of the project and by end-user needs. It is advised that one/several criteria are selected to represent some aspect(s) of the practical feasibility of implementation of each service. Subsequent criteria should represent project aims or underlying end-user needs, therefore capturing a trade-off between what is practically feasible and what the project aims to fulfill. For ease of visualization, two criteria can be chosen, though clearly each service has multiple characteristics and there are likely to be numerous criteria that it would be good to take account of in design decisions. Therefore, it is equally possible to choose more criteria and to explore other visualization options, though there is a trade-off to be considered in terms of usability of the method and the results and method complexity.

The following sub-sections define two decision criteria relevant for electricity access projects for rural villages, and describe methodology for evaluating services against them.

They may not be relevant for all electricity access projects, and users are encouraged to adapt the methods to fit project contexts. Judgment of services against the criteria should be informed by the discussion data gathered in the third step of the SVT, and supplemented by further research. Input should be sought from those familiar with the project context to ensure contextual understandings and interpretations by external investigators are accurate. If it is possible to quantify the extent to which a service complies with a decision criterion, a metric can be defined and calculated for each service; otherwise a categorization of services that represents some kind of ranking in respect to the criterion is advisable.

Near-term Economic Benefit

‘Near-term Economic Benefit’ (NTEB) is an important decision criterion for electricity access projects that seek to provide electricity-enabled services for income-generation, and for communities who identify and rate income generating services highly in the SVT. It encompasses the following considerations:

- Whether the service enables production of a good/service that can be sold (e.g.: chicks hatched using an incubator can be sold, security lights cannot).
- Whether the good/service can be sold outside the village to external markets (e.g.: clothing) or inside the village (e.g.: salon) - and thus whether the good/service takes advantage of external markets, bringing in external income.
- Whether the service complements or enhances an existing service that generates income (e.g.: a TV in a hotel).

Services are categorized into four groups according to their NTEB potential using the decision tree shown in Figure 3.3. Services that provide long-term economic benefit, such as improved health and education, are combined with the enhancement category. This reflects that community services such as health care and education do impact income generation, but probably have more long-term economic impact than near-term impact. The final category distinguishes services that only have social benefit, such as security lighting. Each service should be evaluated taking into account what participants envisage doing with the service, as captured during the SVT and further research if required.

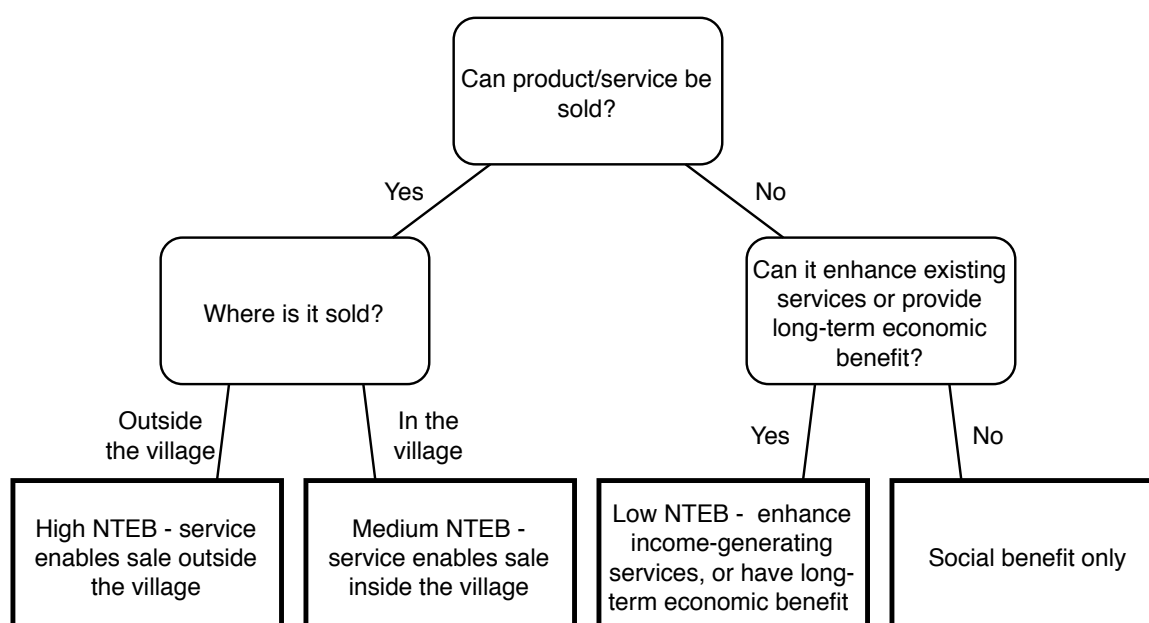


Figure 3.3: Decision tree to categorize Near-term Economic Benefit (NTEB) of services.

Ease of Delivery

‘Ease of Delivery’ (EOD) reflects the feasibility of delivering the services. It can be interpreted to include different factors, and here is defined to include consideration of:

- Power and energy requirements
- Infrastructure requirements
- Other aspects of service complexity, such as the requirement of trained personnel, or a required role by external entities

Firstly, services are separated into power categories as defined by the Multi-tier Framework for energy access (see Figure 1.1 in Chapter 1), shown by the columns of Table 3.1. The power categories are then further partitioned according to services that have an infrastructure requirement, and services that have both an infrastructure requirement and additional complexity factors (the rows of Table 3.1). The result is placement of services within one of the 15 spaces in Table 3.1.

To form EOD categories, the 15 groups defined by Table 3.1 are combined to reflect an approximate indication of how difficult implementation of that service would be. The shading in Table 3.1 shows the group combinations, such that services in each pure power category are combined with services in the previous power category that also have infrastructure requirements, and with services with even lower power requirements that have both

Table 3.1: Template to categorize services according to three components of Ease of Delivery

	Power Categories				
	<50W	50-<200W	200-<800W	0.8-<2 kW	2 kW-
Electric power only					
With infrastructure					
With infrastructure + complexity					

Table 3.2: Summary of the 7 Ease of Delivery categories

#	Groups included
1	<50W
2	50-200W, <50W + infrastructure
3	200-800W, 50-200W + infrastructure, <50W + infrastructure + complexity
4	0.8-2kW, 200-800W + infrastructure, 50-200W + infrastructure + complexity
5	>2 kW, 0.8-2 kW + infrastructure, 200-800W + infrastructure + complexity
6	>2 kW + infrastructure, 0.8-2 kW + infrastructure + complexity
7	>2 kW + infrastructure + complexity

infrastructure requirements and additional complex factors. This method forms 7 EOD categories summarized in Table 3.2.

Combining services in this way synthesizes the three EOD components to give a small number of EOD categories, which helps keep the outcome understandable, while still providing a reasonable distinction between services.

3.2.2.3 Construction of Service Maps

Service maps show two indicators of service popularity: the number of counters the service received during the SVT by the size of each service bubble, the number of focus groups who suggested that service by the color of the service bubble. The axes are decision criteria, and services are arranged according to their categorization with respect to each the criterion.

3.2.3 Step 3: Preliminary System Design

The final step in the SVM method is use of the service maps to select the services the project will deliver. The areas of the service maps that maximize the decision criteria can be considered alongside service popularity to allow designers to make informed decisions in

service selection. The choice of services leads to selection of the products and system architecture that enable those services, and then sizing of the generation, storage and distribution components of the system.

3.3 Case Study: Rural Electrification System Design in Kenya and Bangladesh

SVTs were carried out in the case study communities during the community consultations in 2014–2015. Service maps are presented for both countries, while the resulting electricity system design and implementation details are presented only for Kenya, because design of the system for the Bangladesh context was undertaken by other members of the project team.

3.3.1 Step 1: Data Gathering: SVT Implementation

Three focus groups each comprising 10 participants were carried out in each of the four communities, with separate groups for women, youth (defined as aged 18-30 years, these groups were mixed gender in Kenya, all male in Bangladesh), and elders. Both a household and a community SVT were carried out for each group. For the household SVT, participants used the Free Assignment method to indicate their value judgment, whereas for the community SVT, they used the Ranking method. Each focus group took between 2 and 2.5 hours for both SVTs.

Tables summarizing the content of the discussions in the SVTs for the focus groups carried out in Kenya, in terms of current practices to access that service, and why each service was important, are in Appendix A.3.

3.3.2 Step 2: Data Interpretation: Results

Table 3.3 shows the services categorized according to the extent electricity is the primary component. Appendix A.4 defines each service in terms of what it was assumed to encompass for the purpose of this categorization. Services in the ‘No electricity needed’ column were out of scope of the project.

The remaining electrical services were categorized according to the two decision criteria described in Section 3.2.2. NTEB was relevant because one of the key project objectives was to provide electricity access to enable income generation, and it aligned with a strong theme that emerged from the discussion step of the SVTs. Income generation was one of the

Table 3.3: Services categorized by the extent electricity is a primary component.

Extent of electricity as primary component			
No electricity needed	Electricity and infrastructure	Only requires electricity	
Household: Furniture	Household: Solar powered shower	Household: Torch	Community: Improved roads
Community: Sanitation	Washing machine	Radio	(street lights)
Improved roads	Household water supply	TV	Rice husking
Crop storage	Community: Sports playground	CD/DVD player	Grinding machine
Tractor and pesticides	Church	Fan	Shave cutter
Ambulance	Health center	Lighting	Power tools
Financial support for the elderly	Education	Electric cooking	Both: Security lighting
Alternative mobile phone operator	Pre-primary education	Rice cooker	Incubator
Security wall	Water	Kettle	Welding
Public transport	Community center	Iron	
Fishery	Carpentry	Computer	
Bath	Salon and barber	Printer	
Bank (micro-finance)	Market place	Mobile phone charging	
Improved mobile reception	Garment factory	Fridge	
	Mobile phone charging center	Shaver	
	Both: Irrigation	Sewing machine	
		Hairdryer	

main underlying needs of participants and strongly influenced their suggestions and value judgments. Table 3.4 shows the resulting assignment of services to the NTEB categories split into ‘household’ and ‘community’ groups.

The practical feasibility of system delivery provides an important trade-off to NTEB and service popularity. Table 3.5 shows the classification of services against the three EOD components. Appendix A.4 details the assumptions made about infrastructure requirements and additional complexity for each service. For the Kenyan data, ‘Health center’ and ‘Church’ were separated by village, denoted ‘Health center (L/E)’ and ‘Church (L/E)’, because of differences in existing infrastructure between villages. Lomolo B had no existing health service and no church building, but Echariria had a dispensary building where a nurse saw patients and multiple churches in the community. This removed the infrastructure component of EOD. In the Bangladesh context neither village had infrastructure for health, so the service ‘Health center’ was grouped with the health center for Lomolo B (denoted ‘Health center

Table 3.4: Services categorized according to Near-term Economic Benefit

Near-term Economic Benefit Categories: Low - High			
1 (Low) Social benefit only	2 Enhance other services / long-term economic benefit	3 Sell inside the village	4 (High) Sell outside the village
Household Services			
Solar powered shower	TV	Computer	Welding
Kettle	CD/DVD player	Printer	Irrigation
Torch	Electric cooking	Shaver	Sewing machine
Radio	Rice cooker	Washing machine	Incubator
Fan	Lighting	Fridge	Rice husking
Security lighting		Iron	
Household water supply		Hairdryer	
		Mobile phone charging	
Community Services			
Sports playground	Health center	Shave cutter	Incubator
Security lighting	Education	Market place	Carpentry
Improved roads - (street lights)	Religious school	Salon and barber	Welding
	Pre-primary education	Community center	Grinding machine
	Church	Mobile phone charging center	Power tools
	Water		Garment factory
			Irrigation
			Rice husking

(L, B)' in the tables and service maps).

In addition, in both countries the service 'Education' encompassed improvements to current education buildings and aspirations for new institutions, like secondary schools and polytechnic colleges. To reflect this difference 'Education' is represented twice, as 'improved' and 'new' respectively. When placed on the service map, both reflect the popularity data from the generic 'Education' service, because it was not possible to separate out popularity for 'improved' education and 'new' education.

The classification against the three EOD components produced five EOD categories, shown in Table 3.6, with categories ranging qualitatively from hard to easy and with services split into the 'household' and 'community' groups.

Data from the 6 focus groups in each country were collated to form one household service map and one community service map for each country. Services exist in one of 20 quadrants depending on their evaluation against the decision criteria. To avoid overlapping services,

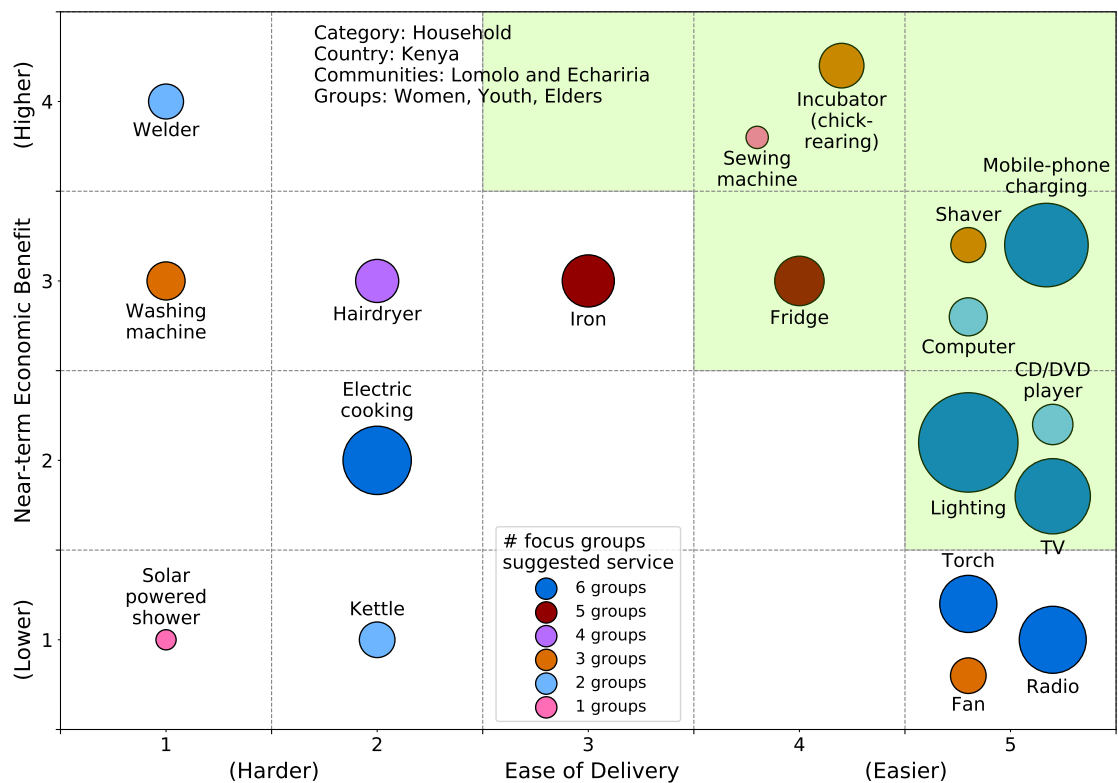
Table 3.5: Services categorized according to three components of Ease of Delivery

	Power Categories				
	<50W	50-<200W	200-<800W	0.8-<2 kW	2 kW-
Electric power only	Torch Radio Lighting Fan Mobile phone charging TV CD/DVD player Computer Printer Shaver	Security lighting (household) Fridge Sewing machine Incubator Security lighting (community) Mobile phone charging center	Rice husker Grinding machine Power tools Improved roads (street lights) Iron Education (improved) Health center (E)	Electric cooking Rice cooker Hairdryer Kettle Church (E)	Welding Shave cutter
With infrastructure (I)		Sports playground Market place	Community center Carpentry Salon and barber	Church (L) Garment factory Water Irrigation Washing machine Household water supply Solar-powered shower	
With infrastructure + complexity (I,C)	Pre-primary education	Health center (L, B) Education (new) Religious school			

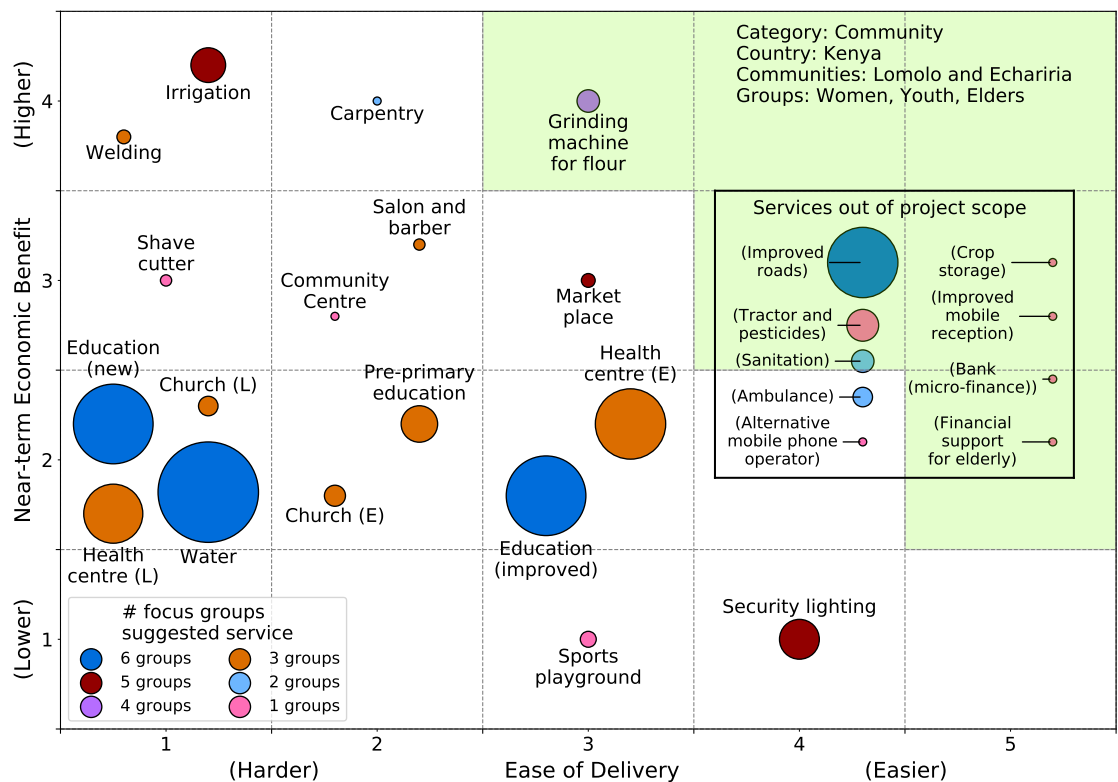
where more than one service exists in a quadrant they are spatially displaced from each other. Figures 3.4 and 3.5 show the service maps for household and community services, for Kenya and Bangladesh respectively.

3.3.3 Step 3: Preliminary Design: Solar Nano-grid

For the Kenyan context, the service maps were used to guide selection of the services the solar nano-grid would provide and build up a system design. The aim was to choose a small number of services, which would be implemented, tested, and developed with in-field learning

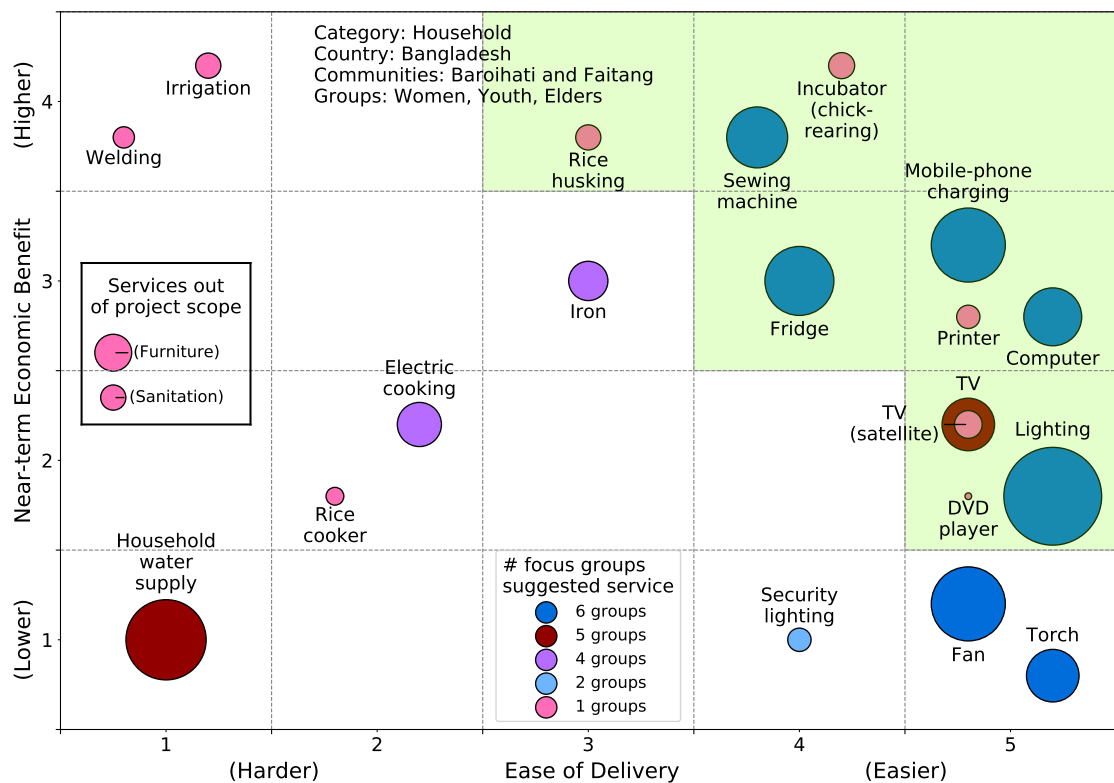


(a) Service map for household services

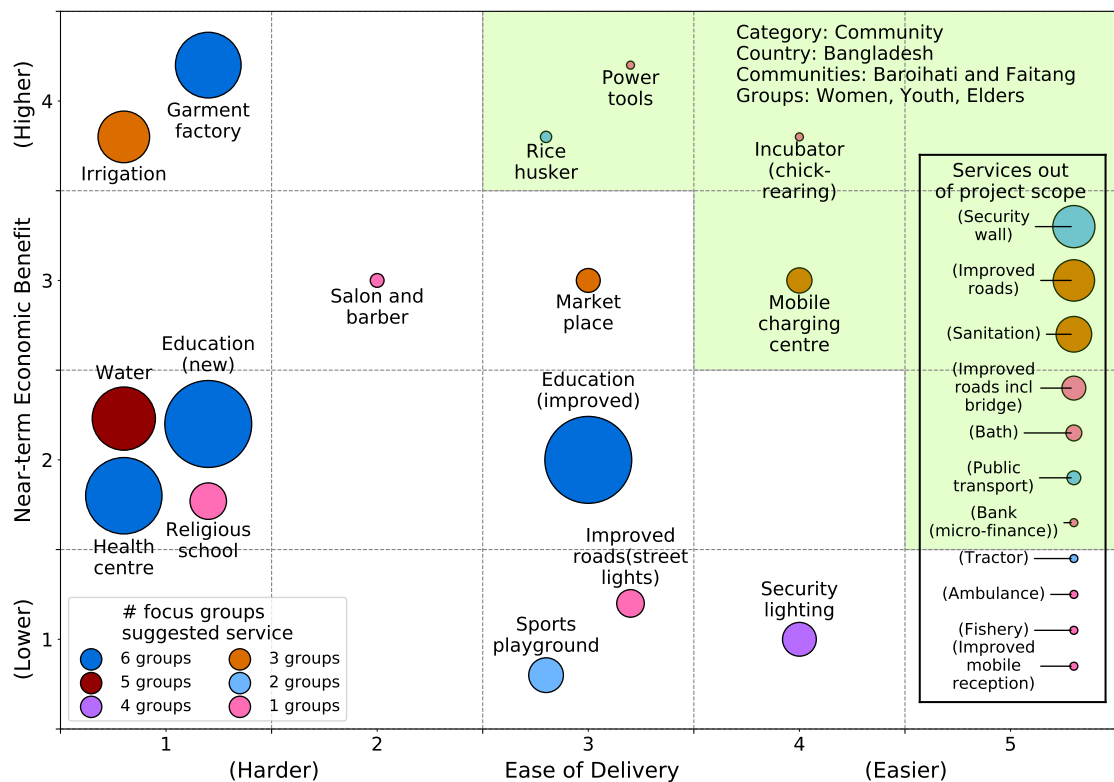


(b) Service map for community services

Figure 3.4: Service maps for a) household and b) community services for Kenya, Lomolo B and Echariria communities



(a) Service map for household services



(b) Service map for community services

Figure 3.5: Service maps for a) household and b) community services for Bangladesh, Faitang and Baroihati communities.

Table 3.6: Services categorized according to Ease of Delivery

Ease of Delivery Categories: Hard - Easy				
1 (Hard) 2kW- 800-2kW, I 200-800W, I, C	2 800-2kW 200-800W, I 50-200W, I, C	3 200-800W 50-200W, I <50W, I, C	4 50-200W <50W, I	5 (Easy) <50W
Household Services				
Welding Shave cutter Irrigation Washing machine Household water supply Solar-powered shower	Electric cooking Rice cooker Hairdryer Kettle	Rice Husker Iron	Fridge Sewing machine Incubator Security lighting (household)	Torch Radio Lighting Fan Mobile phone charging TV CD/DVD player Computer Printer Shaver
Community Services				
Welding Church (L) Garment factory Water Irrigation Health center (L, B) Education (new) Religious school	Church (E) Community center Carpentry Salon and barber Pre-primary education	Rice husker Grinding machine Power tools Improved roads (street lights) Education (improved) Health center (E) Sports playground Market place	Incubator Security lighting (community) Mobile phone charging center	

in partnership with the communities, allowing for refinement and expansion in the future.

The shaded areas in Figures 3.4 and 3.5 show the areas of the service maps that have higher NTEB potential and more favorable EOD. The solar nano-grid was to have a central solar hub, where productive use services would be located, and provide electricity to surrounding houses. Two household services were identified that had high popularity: lighting and mobile phone charging. Using efficient LED bulbs, the power requirement for household lighting is low (<20W), and the SVT discussions informed us that lighting was required in the night rather than the day (duration 3–4 hours). Mobile phone charging is also a low power and energy activity on the scale of a single household. Therefore, rather than put in transmission lines

for a low power and energy requirement, portable batteries were decided on as the means of distribution of electricity from the central hub to the households.

Two productive use services were chosen that would be available to the wider community with the aim of enabling income generation: an incubator for chick-rearing, and a grinding machine to grind maize into flour. Although the incubator was suggested in the Kenyan household SVT and not the community SVT, it was decided to test a centrally located service. An incubator keeps the temperature and humidity constant while the eggs are in development (21 days for chicken eggs), requiring electricity 24 hours a day. Alongside the natural daily and seasonal variation in solar irradiance, this meant electricity storage at the hub was required.

The grinding machine would be the highest power service the nano-grid would support (0.75 kW). It was intended to be an alternative machine to other diesel powered machines in the community available for use for 2–3 hours a day when irradiance is high. The grinding machine takes in maize kernels, already separated from the cob, and grinds them to flour; therefore, a husking machine (0.38kW), which separates kernels from cobs, was also included in the design.

The PV panels and battery system were sized according to irradiation data from near-by weather stations and estimated use of the services. Additional capacity was added given it was a prototype system and demand was unclear. The prototype nano-grids consisted of 3 kW PV panels and 10 kWh battery storage. The household battery charging system consisted of 62 Wh batteries with three 3W LED lights and mobile phone charging connectors. A schematic of the solar nano-grid and the central hub in Lomolo B is shown in Figure 3.6.

3.3.4 Solar Nano-grid Implementation

This section describes the outcome of the installation of two prototype nano-grids in the two case study communities in Kenya in June 2016, where some aspects of the design worked well, and others did not.

When the communities tested the grinding machine, they found that the coarseness of flour it produced was not suitable to their needs. The grinding machines were sourced from a company operating in South-East Asia, who had experience providing agricultural machinery to local communities there. However, in the contexts they worked in, coarse flour was desirable. In Kenya, flour for human consumption must be very fine, and the coarse flour produced by the grinding machines was only suitable for chicken feed. Thus, the

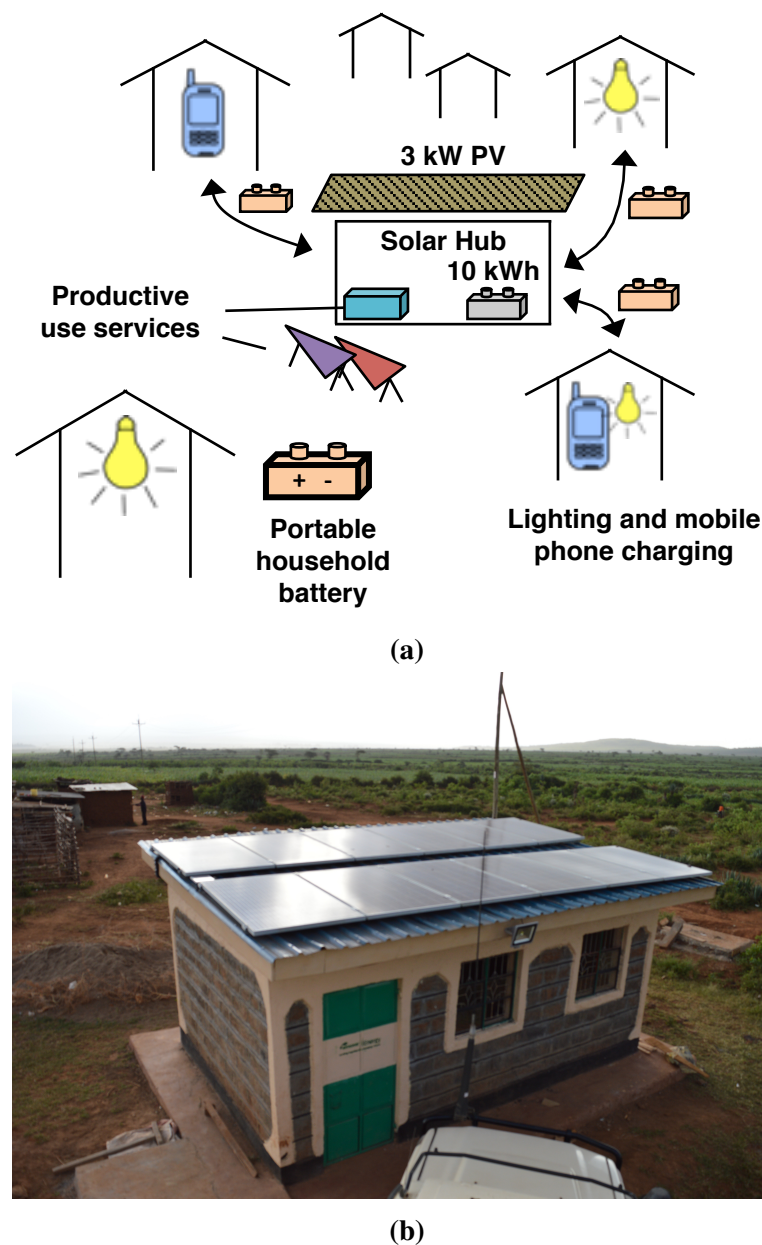


Figure 3.6: (a) Schematic of solar nano-grid prototype and b) photo of Lomolo B hub.

grinding machine did not meet the expectations of the communities in terms of being able to provide flour for human consumption. The husking machine was also unsuitable for these particular villages. This was because farmers accessed a husking machine via a man who had a diesel-powered husker on a motorbike, which he drove to each farm when required.

The incubator was well received and was useful to the communities. The battery charging system was well-used but, alongside technical problems, feedback indicated that end users aspired to household batteries that had larger power and energy capabilities. They made positive reference to the lead-acid batteries available locally that were larger than the smaller

Li-ion packs deployed by the SONG project. While the technical issues were on-going, customers did not pay the planned periodic subscription charge for the batteries. In 2017, larger lead-acid batteries were introduced with a cost-reflective business model such that customers pay monthly until they own the batteries, after which there is a monthly charging payment for use of the charging equipment.

Given that the highest power services were unsuitable, the sizing of the PV panels and hub batteries was easily adequate for demand. Continued development after the June 2016 installation, apart from the introduction of the larger lead-acid batteries for the battery-charging system, was focused on adapting the grinding machine to produce finer flour, adding more incubators, powering computers and printers, and in one community a M-PESA (mobile money) shop was set up in the central hub.

3.4 Discussion and Evaluation

The SVM was developed from user-centered and participatory design approaches, focusing on uncovering and including end-user needs and context in the preliminary design process. In this section, the SVM is evaluated in the extent to which it met the aims of the method during the case study described in Section 3.3, and the flexibility of the method to be appropriate to variants of the situation described in the case study.

The first aim of the SVM is to uncover data on end-user needs, aspirations and surrounding context to an appropriate level of detail for a preliminary system design. In the case study, the SVT data were very useful to inform what services were important to end users, the practical feasibility of implementation and the potential for economic benefit through them. However, the contextual data gathered through the SVT are not sufficient to go beyond a preliminary design to a detailed design and implementation, as exemplified in the case study: contextual factors about grinding flour and current practices around husking maize did not come up in the SVT discussion. However, it is not realistic to expect the SVT to reveal all relevant information required for successful implementation of each service due to time constraints (2.5 hours is already a significant period of time to ask participants to attend for), and the breadth of services covered. The SVT achieves a broader overview of aspirational services and the context around them rather than delving into the detail of a few specific services. Therefore, after the SVM has been used to select candidate services for implementation,

further details on those services should be sought through follow-up visits and communication with the end users, moving from a conceptual design to a detailed technical design. In the case study, in-country partner issues and team miscommunication derailed this process, which also affected the battery charging system. Aspects such as size, weight, charging time, and capacity of batteries should have been more deeply consulted on with end users after the SVM, avoiding design-team assumptions that did not reflect end users' views.

The second aim of the SVM is to enable designers to integrate these data into the design process. Step 2 processes the data into service maps displaying services according to key decision criteria and the popularity data from the SVTs. This is a useful and efficient way to present two aspects of the popularity data alongside criteria that are meaningful for the project, and so guide decision making in terms of what services should be chosen.

In the case study, two key decision criteria were used to construct the service maps and led to a good choice in services selected for the design, despite challenges in the later implementation. 'Ease of Delivery' reflects practical aspects of feasibility of delivering each service, though it does not include a quantified cost of each service. Instead, cost is indirectly included in the three components categorically, as cost of a service will scale with the power requirement, non-electrical infrastructure requirement, and additional complexity factors such as the requirement for trained personnel or interconnection to a national service. For the case study, the definition of EOD in Section 3.2.2.2 was an appropriate level of detail, because the SVM was used as part of the preliminary design process and a detailed cost analysis was conducted after candidate services had been selected from the service maps. However, an adaptable aspect of the SVM is the choice of decision criteria used. Depending on the degree of detailed cost information available and the level of financial detail sought, a quantified cost for implementation of each service could be chosen as a decision criterion. Such a criterion could be defined as broadly as deemed appropriate or relevant (or actually possible with respect to the financial data available) in terms of the aspects of implementation and service life-cycle that should be included in a cost estimate. A challenge with quantifying cost at the early stage of design to which the SVM aims to contribute is the availability of accurate cost data. This is why a categorical definition reflecting practical feasibility of delivery is advantageous to enable rapid assessment at the start of a project. If this route is followed, as demonstrated in the case study, a cost analysis is an important step in the subsequent design process to confirm

the viability of selected services and arrive at a detailed technical design. This is because cost is a key driver for the viability of electricity access systems, individually and at scale.

The second decision criterion described in this research and used in the case study was 'Near-term Economic Benefit'. This was accurately aligned with project and community goals, but did not reflect the other main underlying theme emerging from the SVT discussions - that of a product or service saving the end user time (see Appendix A.3). To better reflect end-user needs in the design process, including more of these underlying themes as decision criteria would have been useful.

The third aim of the SVM is to be fast and effective in the field, to make it easy to integrate into design processes and to address the challenges of remoteness that are common in rural Global South contexts. Although there is some equipment associated with the SVT method, it is not complicated to set up or organize in preparation for the sessions, and the case study showed that a useful depth of information was gathered for both household and community services in 2-2.5 hours. The SVT was effective in including responses from all participants through the quantitative step, because even when participants were shy, or less inclined to participate in the group discussion, their value judgments were still captured and represented in the popularity data. Also, collecting the value quantification data individually avoided group-think or stronger personalities influencing decisions.

However, a barrier to fast and effective implementation of the SVT arose during the case study, as for some groups the concept of 'community' services was not easily understood or relatable. For example, in Echariria, the three spatially dispersed areas corresponded to three different populations who arrived in the area at different times, and who had different socio-economic circumstances. There was understandably less community cohesion across these three areas than there was in Lomolo B where the people had been together (although not in that location) for a long time. Groups in Echariria took longer to understand the concept of community services, gave fewer suggestions, and required more prompting with examples of services than groups in Lomolo B. It is important that local context is considered in preparing the implementation of the SVM, and the method should be reviewed by someone from the local context to determine whether or not it will be suitable and if adjustments need to be made.

The fourth aim of the SVM is to be useful and effective in Global South contexts. The low equipment requirements are an important aspect of this, and in addition there is no requirement for on-site electricity as long as there is enough light to carry out the exercise.

Additionally, in both Kenya and Bangladesh, using pictures (often hand-drawn according to participants' instructions) on the service cards worked well to bridge language barriers among the group and between the facilitators and participants. However, it was more challenging to cope with mixed language abilities in the discussion step of the SVT. Where participants with some English might use English, which was useful for facilitators, it meant other members of the group could not follow the discussion or contribute on the subject. During the case study focus groups, the translator was asked to translate from English to the local language when this occurred to ensure everyone was included, but this slowed down discussions that might have otherwise revealed other useful information.

The SVT method can be easily adapted according to what participants are comfortable with in terms of numeracy and complexity, so it is appropriate to use when literacy barriers are present. During the case study, one group was given 10 crosses/counters rather than 20 for the value quantification Free Assignment process. This made the process easier for some participants who were struggling to count to 20 and who were intimidated by the exercise. The key information can still be captured with fewer counters in the Free Assignment method, as long as the results are scaled up so that all groups are equally represented when the data are analyzed.

For participants with more advanced numeracy skills, extra information could be gathered by introducing time as a variable in the value quantification step, such that participants could either select a service for use in the day or in the night. For a solar electricity project, this would reveal useful information about the time of day services were required, which has implications for the energy storage component of the system. This could be reflected in the service chart by splitting each box on the service chart into two, with one half representing day use, and the other side representing night use. When this was trialled during the case study for one focus group, it quickly became apparent that this was too complex, however, it might be suitable for groups where more time is available, where groups are smaller, or for participants with higher literacy levels.

The SVT process provides a final important function necessary for successful system design. Feedback suggested that participants enjoyed the interactive nature of the SVT process and felt involved in the project. As well as getting useful data to indicate aspirations and the context behind them, the SVT can also help build relationships with the communities

and foster ownership of the project; both of which are important factors for the success of international development projects.

The SVM can be used by a system developer to determine what services should be provided by an electricity system that is in the process of being designed and deployed. In addition to this, the data on aspirational services provides insight into what services will drive demand growth in the future. Knowing what services customers will be wanting to use in the near-future is important if system developers seek to expand or adapt systems over time, in response to changing customer financial capacity to pay and desire for more electrical services. To support this purpose, the SVM could be re-run to monitor how service aspirations change over time, feeding into decision-making of what services would be appropriate to deploy next.

3.5 Summary

This chapter outlines the Service Value Method in response to the question: To build electricity access systems that are suitable for specific contexts, how can we take account of end-user needs, aspirations, and context, in the design process?

The data-gathering step of the SVM uses a simple and pictorial method to gather quantitative and qualitative data about end-user needs, aspirations, and local context. It is adaptable to the capacity for participation of the end users, and as such is suitable for use in the Global South. The second SVM step is a means to interpret the data and integrate it into preliminary design decision making. Service maps are constructed, which summarize the end-user preference data about current and aspirational, and show trade-offs between key decision criteria for the design of the system. The service maps enable the designer to select suitable services, taking into account service popularity with end users, and categorization of services according to decision criteria chosen to reflect project aims and end-user needs. The SVM draws on the service-oriented approach advocated both by energy access practitioners and the sustainable product-service systems approach.

This chapter outlined the design of the solar nano-grids for two communities in Kenya, which is the case study for this thesis. The remaining chapters focus on the use of the battery-charging system in the solar nano-grids; the use of the productive use services at the central hub is out of the scope of this research.

How can the accuracy of time series electricity data collected from multiple sensor nodes be improved when real-time time synchronization is not possible, and there is limited time calibration data available?

4

Processing Electricity Data for Rural Nano-grid

Contents

4.1	Introduction	79
4.2	Time Drift and Correction in Remote Sensing	87
4.3	Methodology	89
4.3.1	Definitions of Drift Rate and Time Correction	90
4.3.2	Estimating Drift Rates for Time Correction	91
4.4	Case Study	95
4.4.1	Available Data: Measured Drift Rates	95
4.4.2	Drift Rate Estimation	97
4.4.3	Time Correction	100
4.4.4	Accuracy of Time Correction	100
4.4.5	Final Result: Household Electricity Profiles	107
4.5	Discussion	108
4.5.1	Quality of Electricity Profiles	108
4.5.2	Use of the Data in Further Work	110
4.6	Summary	114

4.1 Introduction

The previous chapter outlined the Service Value Method, which was used to include socio-cultural and contextual factors in the design of a solar nano-grid system. The design and implementation of two solar nano-grids in Kenya was described. The solar nano-grids were designed to meet household electricity needs via a battery-charging system, which enabled

Table 4.1: Key parameters stored by microcontrollers in household batteries

Parameter	Description
Time stamp	Format ‘Ddd mmm ss hh:mm:ss yyyy’
Output voltage	Voltage at the 12V port
Output current	Current at the 12V port
USB current	Current at the USB port
Current controller	The duty cycle of the current controller on the 12V port
Battery state	The state of the battery indicates what is happening to the battery. Three states were used in later processing stages.
State of charge (SOC)	Approximate indication of SOC of the battery

lighting and mobile phone charging services. To address the lack of electricity use data for electricity access systems, the household batteries in the battery-charging systems were deployed with monitoring capabilities, tracking household electricity use over the trial period. This chapter describes how the raw data set gathered by the household batteries was processed to give useful electricity profiles for each household, which can then be used to support design and research of off-grid electricity access systems. A specific contribution is made to address the inaccurate timestamps of the electricity data caused by drift of the sensor clocks, in the form of a method to correct the timestamps post data collection.

The household batteries (manufactured by Brill Power) deployed are shown in Figure 4.1, with labelling of the hardware inside it and the connection ports. When the battery was switched on, a microcontroller periodically recorded parameters to an SD card alongside a timestamp from the microcontroller real-time clock (RTC) at intervals of every 4–8 seconds. When the battery was switched off, the microcontroller was also switched off, and the RTC was powered by a back-up battery on the underside of the microcontroller in the photo. The data were stored in comma-separated values (csv) files, with a new file made for each hour of data collection. 15 parameters were stored in each reading, of which 7 are important for subsequent data preparation and analysis and are listed in Table 4.1.

The two prototype systems were installed at the end of June 2016, and the trial period ran until the end of November, at which point battery data sets were collected by reading the SD cards during site visits. During the trial, 52 regular users across both sites used the household batteries. There were reliability issues with the batteries, which meant that users swapped batteries when repairs were necessary, and the number of users dropped throughout the trial in general. In addition, some users used two batteries at a time out of preference. As a consequence, data for 137 batteries were gathered.



Figure 4.1: Photo of household battery in the case study, external view (top) and internal hardware (bottom).

The first stage of data analysis in a study such as this one requires raw data to be cleaned and transformed so as to be usable, which can be a lengthy and complex process while ensuring data integrity is maintained [132]. Without processing the data to a usable and accurate format, it may not be possible to use it for analysis, or it may not yield accurate results in analysis. The steps taken to process the data are explained and discussed here for the reference of others attempting similar operations.

There were 13 steps in the process to construct electricity profiles for each household using the systems. Figure 4.2 shows a comparison of heat maps of the raw data for a single battery (a power profile with data points every 4-120 seconds), and the final result of the data processing: a power profile on a half-hourly basis attributable to a single household. Months are marked on the x-axis, with dotted grid-lines showing weeks, and time is on the y-axis, such that each day runs from the bottom of the map to the top. The positive power values in shades of red indicate where the battery is powering a load, and the negative power values in shades of blue

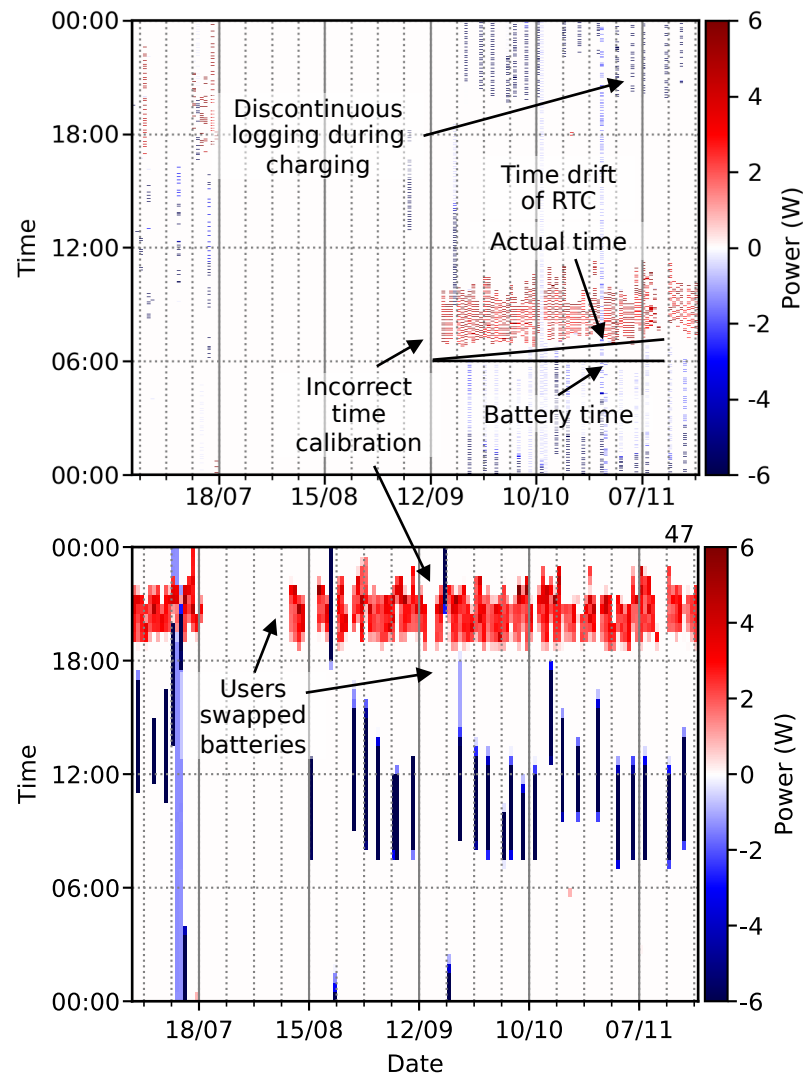


Figure 4.2: Comparison of raw and processed electricity data using a heat map of power on the 12V battery port. Raw data for a particular battery (top) and processed data for household 47, to whom the majority of data from the battery are attributed (bottom). Four of 13 issues shown, with the time drift shown not to scale.

indicate where it is being charged. The scale for the heat maps is $\pm 6\text{W}$, even though the power during charging is more negative than -10W . This is to achieve an adequate distinction in visualization of the positive power values, where the battery is being used to power loads. Heat maps are used to portray the usage as they make patterns easy to see and interpret.

On Figure 4.2, four of issues dealt with in the data cleaning process are shown. The most significant challenge was that without a real-time time synchronization procedure during the data collection period, the RTCs on the microcontrollers in the batteries drifted from the actual time. As shown on Figure 4.2 (not to scale), the battery time became increasingly offset from the actual time. The time series data therefore needed to be corrected post-trial.

Also shown is an incorrect time calibration within the monitoring period, so usage is logged during the morning (between 06:00 and 12:00 on the upper heat map) rather than evening (corrected to between 18:00 and 00:00 on the lower heat map). In addition, logging of data sometimes stopped during charging, giving a poor reflection of charging time and energy, and batteries were swapped between households. On Figure 4.2, two points at which this household swapped from one battery to another are annotated. Table 4.2 summarizes these issues along with nine others not annotated on Figure 4.2, describing the issue, how it was fixed, and where further information can be found (if applicable).

For the issue of varying time drift across the batteries (the first issue in Table 4.2), a significant contribution was made in terms of a robust methodology for post-monitoring time correction of data from distributed sensor nodes, when real-time wireless communication was not possible. This is presented in the main body of this chapter. Where more information about other processing steps might be useful to other people in a similar situation, this is provided in an appendix as referenced in Table 4.2.

The 137 battery data sets were converted to 76 user profiles. Twenty-four households used two batteries for some duration of the monitoring period. Combination of the user profiles resulted in 51 electricity profiles with half-hour time resolution attributable to households out of the 52 regular household users (for one of the 24 households using two batteries, it was not possible to confidently combine data as described in Step 11 in Table 4.2). The electricity profiles constructed in this chapter will be utilized throughout the remaining chapters of this thesis.

This chapter is structured to first introduce the issue of time drift in remote sensing, and describe the time drift observed for the household batteries in the solar nano-grids. Section 4.3 then outlines a methodology for estimating drift rates for a subset of sensor nodes for which there is no measured battery and actual time data, from existing data for other sensor nodes. This method can be used in situations where wireless communication between sensor nodes has not occurred or has been disrupted, to correct data timestamps post-trial. The methodology is demonstrated in Section 4.4 to correct the timestamps for the electricity data gathered by household batteries in the solar nano-grid. A final discussion outlines how accurate the time correction process was, the integrity of the final electricity profiles, and discusses issues to be aware of when using the data in general, for new contexts, and for other electricity systems.

Table 4.2: Summary of challenges and steps to process electricity data into household profiles

Description of issue or step	Fix	Reference
1. Battery RTCs drifted by different and sometimes unknown rates		
The household battery RTCs drifted from the actual time with different drift rates, so data points were recorded against incorrect timestamps. For a minority of batteries, recorded time points of battery and actual times enabled drift rate measurements that could be used for time correction. For 113 of 137 batteries, time points were missing or partially missing, so it was not possible to measure drift rates required to correct the timestamps.	A methodology to estimate drift rates for batteries without drift rate measurements, from existing drift rate measurements for other batteries. This is a process of statistical tests to determine the relationship between drift rates across periods and across the sample.	Section 4.2
2. Unrecorded time calibrations within monitoring period		
For some batteries, visiting personnel and technicians corrected the battery time to the actual time during the monitoring period, without recording this externally. This disrupted the time stream and needed to be taken account of during the time correction. In some cases, the time was set incorrectly.	The battery ‘state’ was recorded alongside every timestamp. This indicated when the time was being re-calibrated. An automatic check of battery state was implemented to find time calibrations, and these were included in the time correction calculations. Incorrectly calibrated batteries were identified and the correct time inferred from heatmaps.	Appendix B Section B.1
3. Battery RTCs were reset to 1970		
Forty-nine batteries were reset to ‘0’ in UNIX time during repairs, logging data against 1970 timestamps for significant periods of time. Where this happened multiple times for the same battery, new data were recorded after old data in the same csv file.	Where possible, layered data readings were identified and removed to leave clean data. For five batteries, readings of battery time and actual time meant these could be calibrated to 2016. A further four were calibrated under assumptions.	Appendix B Section B.2
4. Failed data entries and insignificant data		
Failed data entries caused recordings of ASCII strings, duplications of entries and sets of readings that were not monotonic in time. Some batteries had very short data sets.	Failed data entries were detected when numerical operations failed and were removed. An automatic check for monotonicity was run, and non-monotonic sections were deleted if they were duplicates, or correctly ordered if not. Batteries with insignificant data (< five days) were ignored.	N/A
Continued on next page		

Continuation of Table 4.2

Description of issue or step	Fix	Reference
5. Current calibration		
The current measured on the 12V port was internal battery current I_b , rather than load current I . The current measured on the USB port is current going into the 5V converter, $I_{USB,b}$, rather than load USB current, I_{USB} .	Calibrate 12V port current using equation (4.1) with the duty cycle of the current controller, D_{cc} . Calibrate USB current using equation (4.2), where V_{out} is the output voltage on the 12V port. $I = I_b \times (1 - D_{cc}) \quad (4.1)$ $I_{USB} = I_{USB,b} \times \frac{V_{out}}{5.0} \quad (4.2)$	N/A
6. Microcontrollers logged data with inconsistent frequencies		
The microcontrollers in the batteries logged readings at variable time intervals, and analysis required data to have regular timestamps.	Battery profiles were re-indexed on a per second basis. An algorithm was used to interpret whether long gaps between logged values was due to a battery switch-off, or another delay, and forward fill values accordingly.	Appendix B Section B.3
7. SD cards swapped between batteries		
Five batteries had SD cards in them that did not belong to them. At an undocumented occasion, the old SD card had been removed and a different SD card inserted. This meant some batteries carried sections of data that were not collected on that battery.	The time of the swap was estimated from records kept by the technicians about battery repairs. The battery data set was then split and assigned to the correct battery number.	N/A
8. Batteries were swapped between users		
Batteries were swapped between users, usually if a battery required a repair or stopped working. This meant most batteries had data for multiple users. User-specific profiles were required.	The technicians in the communities kept records when batteries were assigned to users after a repair or a check, and these were used to form 76 profiles across 52 users.	N/A
9. Addressing logging disruptions during charging		
Batteries sometimes stopped logging during charging, giving inaccurate indications of charging energy and behaviour.	An automatic check to find discontinuities in the SOC of the battery was used to find charging events where logging stopped. Typical charging profiles were estimated from data for complete charging events, and superimposed onto the discontinuities.	Appendix B Section B.4
Continued on next page		

Continuation of Table 4.2											
Description of issue or step	Fix	Reference									
10. Power and duration of use limits											
For analysis, days that reflect actual use of the battery system are of interest. There were some days where batteries were left on without connection to a load, were left at the charging hub fully charged, or flicked off and on briefly. These do not reflect actual use to power loads, and so these days should be removed so they are not included in analysis.	Power and duration limits were explored with demand and charging profiles separately. The power limit and duration limit were increased and the number of days eliminated was tracked. Judicious values were chosen to remove data that did not reflect actual use, shown in the inset table: <table> <tr> <th></th><th>Demand</th><th>Charging</th></tr> <tr> <td>Power (W)</td><td>1.5</td><td>-1</td></tr> <tr> <td>Duration (minutes)</td><td>10</td><td>10</td></tr> </table>		Demand	Charging	Power (W)	1.5	-1	Duration (minutes)	10	10	Appendix B Section B.5
	Demand	Charging									
Power (W)	1.5	-1									
Duration (minutes)	10	10									
11. Combining user profiles into household profiles											
Twenty-four households used two batteries at the same time for some portion of the monitoring period. In each case, data from the two batteries needed to be combined to make a household specific electricity profile.	User profiles were combined so that each household profile represented the actual electricity use of a household. For example, if data for just one of two batteries existed, then the data set from the other was also discarded. If one of two batteries was under repair, such that usage was transferred wholly onto the other battery, the data set was kept.	N/A									
12. Half-hourly average electricity profiles											
	Each household profile was converted to a half-hourly average power profile.	Section 4.4.5									
13. Testing loading conditions to ensure ‘normal’ use data											
Electricity use can differ between loading conditions, such as seasonal variations and days of the week. Abnormal and uncommon loading conditions can cause abnormal electricity use. If different loading conditions correspond to different electricity use, common loading conditions must be isolated and used for the analysis. Identifying and removing abnormal data that correspond to abnormal loading conditions is also important, as the objective is to analyze typical electricity use for each household.	The data set was investigated to determine if electricity use differed by month and day of the week. Representative load profiles were formed for each loading condition and compared, and the average daily electricity was plotted for the monitoring period. The data set was restricted between July 1st – 21st November, as June had very high electricity use and data for December only existed for one person. All days of the week were retained.	Appendix B Section B.6									

4.2 Time Drift and Correction in Remote Sensing

The RTCs on the microcontrollers in the household batteries drifted from the actual time, because there was not a real-time time synchronization procedure throughout this study's data collection period. The time series data therefore needed to be corrected post-trial to give electricity profiles that accurately indicate the time of electricity use in these communities. At certain points throughout the trial, the battery time and actual time were logged, and could be used to calculate rates of time drift (drift rates) of the RTCs for time correction. However, these measured time points only existed for a sample of batteries. The issue was compounded because the drift rates were observed to vary across the microcontrollers in the batteries. They were also an order of magnitude higher than expected from specification documents about the equipment used. To address this, a methodology was developed to estimate drift rates for sensors without measured time point data, from a sample of sensors that did have data.

To put this problem in perspective, the acquisition of data by remote and distributed sensing is becoming increasingly common. Examples include monitoring of environmental phenomena such as glaciers [133], monitoring use of services and technology such as water pumps in Zambia [134], and monitoring electricity use in various electricity systems, from grid-connected smart meters to SHSs in the rural Global South [135]. Microcontrollers are typically used in these applications for data logging and other control aspects of the sensing equipment. RTCs in microcontrollers sourced with crystal-quartz oscillators have accuracies dependent on the quartz quality, age, and other environmental conditions [136]. It is a known issue that a clock will drift away from the ideal time, and that the rate of drift will change over time. Management of this drift in monitoring equipment through clock synchronization is particularly important when there are multiple sensing nodes in the system, in multiple products or locations, which must have a common time reference for meaningful data analysis.

Clock synchronization can occur in real-time if distributed sensors are arranged in wireless sensor networks (WSNs); wireless ad hoc networks of distributed sensors gathering data and transporting it to a central location. Frequent periodic synchronization is the best approach to mitigate the variable rate of drift. Research into WSNs focuses on making low-cost, low-power, and reliable systems, and clock synchronization is an important research topic for data accuracy, power management, and transmission scheduling [137]. A good expansion on the challenges and solutions to clock synchronization is found in [137]. [138] is an example

of a research project where time synchronization is a key component in development of the SensorScope; a reliable, low-cost and low-energy WSN for monitoring of environmental data.

The methodology presented here for post-monitoring time correction of time series data is relevant to situations where data are gathered in a distributed remote sensor system, but there is no wireless communication between the nodes or host. This means that real-time clock synchronization is not possible, as in the case for the household batteries in the solar nano-grid deployments. In remote sensor nodes, communication is an energy-intensive activity compared to logging data. Removing the requirement for wireless communication would reduce the energy requirements of the device, enabling a smaller battery to be provided. If regular access to the sensors occurs due to maintenance visits, during which data could also be collected, this is a viable alternative to a WSN.

For the household batteries in this study, the measured drift rates are of the order of 300 ppm in terms of the accuracy of an oscillating crystal. Over four months, this results in an offset from the actual time of 52 minutes. To give the opportunity for detailed analysis of usage profiles later, an accuracy of ± 10 minutes compared to the actual time is acceptable. The offset caused by the time drift is therefore large enough for it to be worthwhile to correct it.

The microcontroller in the batteries was a MK64FN1M0VLL12 on a FRDM K-64F mbed development board [139], and the accuracy of the crystal used in the 32 kHz oscillator that sources the RTC on the microcontroller is reported as ± 20 ppm at 25°C [140]. The measured drift rates are larger than the reported crystal accuracy by an order of magnitude, and all correspond to a speeding up of the RTC, rather than resembling a uniform distribution centred on 0 as suggested by the data sheet. The reported temperature coefficients that affect crystal accuracy are also too small to account for this large difference in drift rates. This suggests that the measured drift rates are not explained by the crystal accuracy, or by the reported temperature behaviour, and that there is something else that is a dominating factor causing these higher drift rates. As it is not possible to know what this factor is, rather than attributing the drift rates to a particular variable, the methodology presented here makes inferences about drift rates solely from the measured data.

4.3 Methodology

The methodology is applicable when readings of sensor time and actual time are available for some sensor nodes, but not all. Readings of sensor time and actual time are referred to as ‘timepoints’ hereafter. Two timepoints are sufficient to calculate a rate of drift of the RTC’s in the sensors, applicable for the time period in between them. This can then be used to correct the timestamps for data collected between the two timepoints. If, for some sensors, timepoints were not measured at either end of the monitoring period, it is not possible to calculate a drift rate to correct the data. This method uses two statistical tests to estimate the drift rates for sensors that are missing time point data.

The method is applicable when time points have been measured three times throughout the monitoring period, because of the assumptions made about the drift rates that cannot be observed due to missing data (detailed throughout this section and explained further below). The three time points must be measured at the beginning (time point (1)), at some point during the monitoring period (time point (2)), and at the end (time point (3)). From three time points, two drift rates can be calculated: the first relevant for the data collected between time points (1) and (2), and the second relevant for the data collected between time points (2) and (3). Drift rates calculated from measured time points are referred to as ‘measured drift rates’, to distinguish from drift rates estimated through this methodology. All sensor nodes must fit into one of the following three categories, according to the availability of time point data, and hence the potential to calculate drift rates:

1. Group A: Measured drift rates available for both periods.
2. Group B: Measured drift rate available for first period, but not the second.
3. Group C: No measured drift rate data available.

Correction of timestamps for Group A is simple because the measured drift rates can be used. For Group B and C, drift rates need to be estimated from existing measured data.

The constraints of the method are that it is applicable for when there are three time points, and so two periods for which the drift rate has been measured (for a sample of sensors). This enables the relationship between the drift rate in the first time period, and the drift rate in the second, to be investigated as part of the sequence of tests to estimate the missing drift rates. Extension of the methodology to be applicable for when there are more than two time

periods for which there is measured drift rate data, in the various permutations of missing data that may arise, could draw on further models in the statistics literature. It is beyond the scope of this research to treat more complex situations beyond the scenario offered by the time drift of the household batteries in the solar nano-grids.

It is also necessary that if sensors are missing a measured drift rate for one of two time periods, they are missing it from the second period, and not the first. This is because the type of statistical test used to infer the relationship between drift rates from the first period and the second period implies that they are sequential in time, such that the drift rate to be estimated occurs after the measured drift rate.

A reduction in complexity, such that just two time points either end of the monitoring period mean that just one measured drift rate is available for a sample of sensors, presents a simple case for drift rate estimation. Missing drift rates would simply be estimated by the sample mean.

This section begins with definitions of drift rate and the time correction process, and then outlines the method for estimating drift rates for sensor nodes which do not have sufficient measured time point data. The method is explained from the perspective of correcting timestamps in the data logging of the batteries.

4.3.1 Definitions of Drift Rate and Time Correction

The drift rate over a period of time can be calculated if time readings are collected from the battery and from a reliable source at either end of the time period of interest. Let T denote time readings according to a battery, and t denote the corresponding actual time reading. Using a to denote the time readings taken at the start of the time period, and b to denote the readings at the end, equation (4.3) defines the drift rate $d_{a:b}$ over the time period as:

$$d_{a:b} = \frac{(T_b - t_b) - (T_a - t_a)}{t_b - t_a} \quad (4.3)$$

The numerator in this equation is the amount of time the battery has drifted over the time period $(t_b - t_a)$, calculated from the difference between the offsets between the battery and the actual time at either end of the period. This is divided by the time period length to give the drift rate; the amount of time that the battery gained or lost per second over that time period. Equation (4.3) is equivalent to:

$$d_{a:b} = \frac{T_b - T_a}{t_b - t_a} - 1 \quad (4.4)$$

An alternative definition of the drift rate is:

$$d_{a:b} = \frac{T_b - T_a}{t_b - t_a} \quad (4.5)$$

The second definition in equation (4.5) defines the drift rate as the ratio of the time that has passed as measured by the battery compared to the actual time. Compared to the first definition given by equation (4.3), and best illustrated by the simplified form given by equation (4.4), the difference is a constant shift of 1. Either definition can be used with the methodology and the resulting corrected battery times would be the same. In this research the first definition (equation (4.3)) is preferred so that the drift rate $d_{a:b}$ equals 0 when there is no drift of the battery RTC.

The methodology is presented assuming the monitoring period is split into two periods, with $d_{1:2}$ referring to the drift rate between time point 1 and 2, and $d_{2:3}$ referring to the drift rate between time points 2 and 3.

Let T_i be a battery time that requires correcting, and t_i the unknown actual time that T_i requires correction to. Let T_a and t_a continue to represent a recorded battery time reading and actual time from earlier in the time period, and $d_{a:b}$ be the relevant drift rate estimate. Once an estimate of the drift rate is found, equation (4.3) can be rearranged to make t_i the subject, giving equation (4.6) which corrects T_i to t_i .

$$t_i = \frac{(T_i - T_a)}{d_{a:b} + 1} + t_a \quad (4.6)$$

4.3.2 Estimating Drift Rates for Time Correction

For batteries that have measured drift rates applicable to a certain time period, equation (4.6) can be used to correct the battery time readings. For Group B batteries, with measured drift rate available for just one of two periods, and Group C batteries, with no measured drift rate data, an estimation of drift rate is required.

The method consists of two hypothesis tests on the measured data to provide insight into the drift rate behaviour. Let $d_{1:2,i}$ be the drift rate calculated between time points (1) and (2) for battery i , and $d_{2:3,i}$ be the drift rate calculated between time points (2) and (3). The following assumptions are made:

- Each observation of drift rate $d_{1:2,i}$ is a realization of a random variable, which has distribution $D_{1:2}$:

$$D_{1:2} \sim N(\mu_{1:2}, \sigma_{1:2}^2) \quad (4.7)$$

- Each observation of drift rate $d_{2:3,i}$ is a realization of a random variable, which has a distribution $D_{2:3}$:

$$D_{2:3} \sim N(\mu_{2:3}, \sigma_{2:3}^2) \quad (4.8)$$

Through these assumptions we allow the means and variances of the two underlying populations to differ, but to facilitate hypothesis testing it is assumed that the observations are from normal distributions.

Figure 4.3 sets out the steps in the method, including how the results are interpreted to indicate what drift rates should be used for Group B and C batteries. The first hypothesis test is a Student's t-test to investigate whether the two means of the distributions, $\mu_{1:2}$ and $\mu_{2:3}$, are equal. The null hypothesis $H_{0,1}$: $\mu_{1:2} = \mu_{2:3}$, the test is two-tailed, and we use a 5% significance level for the test.

For $n_{1:2}$ observations of $D_{1:2}$, and $n_{2:3}$ observations of $D_{2:3}$, the sample mean for observations $d_{1:2,i}$ is:

$$\bar{x}_{1:2} = \frac{\sum_{i=1}^{n_{1:2}} d_{1:2,i}}{n_{1:2}} \quad (4.9)$$

and $\bar{x}_{2:3}$ is calculated similarly. The sample variances $s_{1:2}^2$ and $s_{2:3}^2$ are calculated according to equation (4.10), with the equivalent formula for $s_{2:3}^2$:

$$s_{1:2}^2 = \frac{\sum_{i=1}^{n_{1:2}} (d_{1:2,i} - \bar{x}_{1:2})^2}{n_{1:2} - 1} \quad (4.10)$$

The test statistic is given by equation (4.11).

$$t_s = \frac{(\bar{x}_{1:2} - \bar{x}_{2:3}) - (\mu_{1:2} - \mu_{2:3})}{\frac{s_{1:2}^2}{n_{1:2}} + \frac{s_{2:3}^2}{n_{2:3}}} \quad (4.11)$$

This is compared with Student's t-distribution with degrees of freedom ν given by equation (4.12).

$$\nu = \frac{\left(\frac{s_{1:2}^2}{n_{1:2}} + \frac{s_{2:3}^2}{n_{2:3}}\right)^2}{\frac{s_{1:2}^4}{n_{1:2}^2(n_{1:2}-1)} + \frac{s_{2:3}^4}{n_{2:3}^2(n_{2:3}-1)}} \quad (4.12)$$

The outcome informs what drift rate is used for Group C batteries, which have no measured drift rate data. If the null hypothesis that the means are equal cannot be rejected, it is assumed

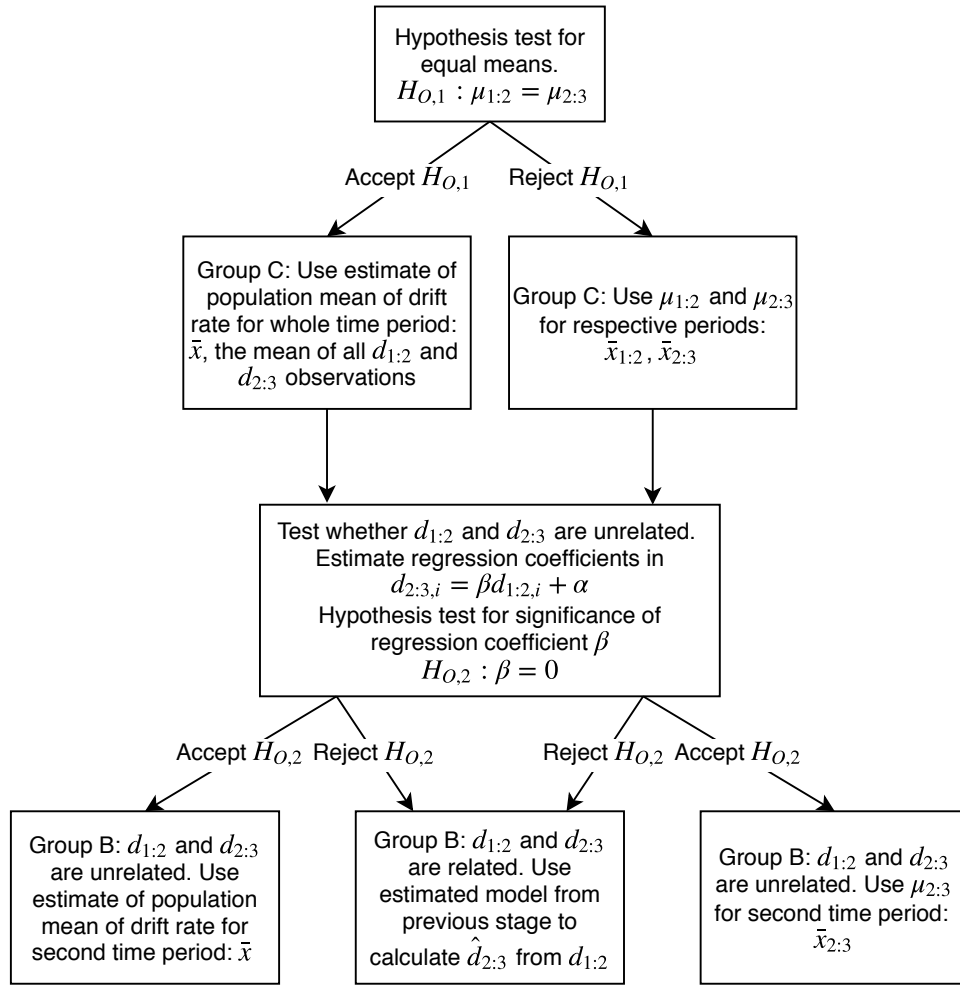


Figure 4.3: Flow diagram of method to infer drift rates for batteries with missing time point data.

that the population means $\mu_{1:2}$ and $\mu_{2:3}$ are the same. This means that the drift rate for Group C batteries should be estimated by using the population mean value μ for the whole monitoring period. The population mean can be estimated by the sample mean of all the observations $d_{1:2}$ and $d_{2:3}$, according to equation (4.13):

$$\bar{x} = \frac{\sum_{i=1}^{n_{1:2}} d_{1:2,i} + \sum_{i=1}^{n_{2:3}} d_{2:3,i}}{n_{1:2} + n_{2:3}} \quad (4.13)$$

If the null hypothesis can be rejected, it is assumed that the population means $\mu_{1:2}$ and $\mu_{2:3}$ are different, and an estimate for each according to (4.9) should be used for time correction of Group C batteries for the respective time periods.

The second step is to test whether the $d_{1:2}$ and $d_{2:3}$ drift rates are unrelated, which determines the drift rate to be used for Group B batteries that lack drift rate data for the second time period of monitoring. It seems reasonable to suppose that $d_{1:2}$ and $d_{2:3}$ may be correlated, such that batteries with a large $d_{1:2}$ would likely have a larger than average

$d_{2:3}$. If there is a significant relationship between $d_{1:2}$ and $d_{2:3}$, unknown drift rates could be predicted from this relationship. This is tested for the n batteries for which observations of both $d_{1:2}$ and $d_{2:3}$ exist, by fitting a linear regression model to the data. More complicated relationships could be allowed in principle, however this method was developed for data with a small sample size, and by considering a linear regression model we essentially consider whether the covariance is non-zero.

It is assumed that there is a first order relationship between $d_{1:2}$ and $d_{2:3}$ according to:

$$d_{2:3} = \beta d_{1:2} + \alpha \quad (4.14)$$

For each pair of observations for $i = 0, \dots, n$, the underlying relationship between $d_{1:2,i}$ and $d_{2:3,i}$ is described by equation (4.15), which includes an error term e_i , capturing the deviation of the observed values from the model in equation (4.14).

$$d_{2:3,i} = \beta d_{1:2,i} + \alpha + e_i \quad (4.15)$$

An ordinary least squares approach minimizing the sum of the squared residuals $\hat{e}_i$ is used to estimate coefficients β and α as $\hat{\beta}$ and $\hat{\alpha}$. An assumption of homoscedasticity is made, such that the variances of the error terms in the regression are assumed to be the same.

A Student's t-test is used to investigate the significance of the regression coefficient β of $d_{2:3}$ on $d_{1:2}$. The null hypothesis $H_{0,2}: \beta = 0$, and the alternative hypothesis $H_{1,2}: \beta \neq 0$. The test is two-tailed with 5% significance. The test statistic is given by equation (4.16), where $\hat{\sigma}_{\hat{\beta}}$ is the estimated standard error of the estimate $\hat{\beta}$.

$$t_s = \frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}}} \quad (4.16)$$

This is compared with Student's t-distribution with degrees of freedom:

$$v = n - 2 \quad (4.17)$$

The null hypothesis is either affirmed or rejected depending on the outcome. If the null hypothesis is rejected, we assume that the relationship between $d_{1:2,i}$ and $d_{2:3,i}$ can be described by equation (4.14) with the estimates of the regression coefficients, and missing drift rates for Group B can be estimated. If the null hypothesis is accepted, there is not a significant relationship between drift rates of the two time periods. As Figure 4.3 shows, depending on the outcome of the first hypothesis test, the drift rate is estimated as the overall population mean, or the population mean specific to the time period in question.

Table 4.3: Time points recorded throughout solar nano-grid trial

Time point	Approximate Time and Stage of the Trial	Symbol
Time point (1)	March 2016, during battery manufacture when the RTC was first calibrated	Battery time T_1 Actual time t_1 ($T_1 = t_1$)
Time point (2)	June/July 2016, during deployment of batteries into the communities	Battery time T_2 Actual time t_2
Time point (3)	November 2016, during data collection visit	Battery time T_3 Actual time t_3

Table 4.4: Number of batteries with measured drift rates for each time period

	$d_{1:2}$	$d_{2:3}$	$d_{1:3}$
Number batteries	115	20	2
Number batteries with outliers removed	111	18	2

4.4 Case Study

This section outlines the use of the method to estimate drift rates for the 137 household batteries in the solar nano-grids and correct timestamps to actual time.

4.4.1 Available Data: Measured Drift Rates

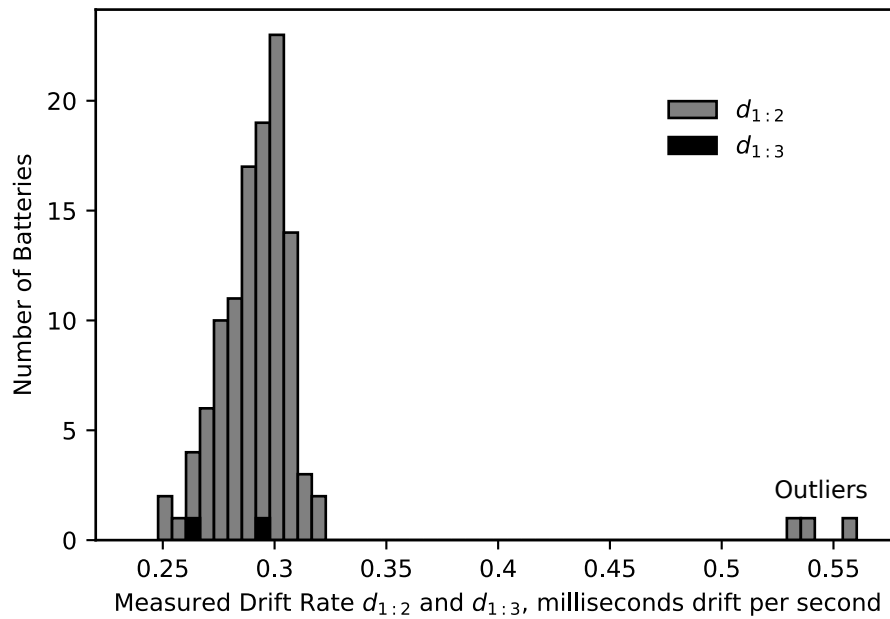
At three points during the trial, the specific time and date of the RTC of a subset of batteries, and the corresponding actual time, was recorded. The time points are given in Table 4.3 along with an indication of when and what stage in the project they were recorded.

Given the time point data, the number of batteries for which drift rate can be calculated is shown in Table 4.4. The majority of batteries have a measured drift rate for $d_{1:2}$, and far fewer have a measured drift rate for $d_{2:3}$. Two batteries have time points (1) and (3), so a drift can be calculated for the whole period. Table 4.5 shows the number of batteries in the Groups A, B, and C as specified by the availability of drift rate data.

Both Tables 4.4 and 4.5 reference outlier data points. Four batteries were labelled as outliers because they were abnormal in comparison to the rest of the observations. Three of these were observations of $d_{1:2}$ that were abnormally high. Figure 4.4 shows a histogram of the 115 drift rates calculated between time points (1) and (2), $d_{1:2}$, and the two drift rates calculated between time points (1) and (3), $d_{1:3}$, showing three abnormally high drift rates. Excluding these high values, the mean and standard deviation of the sample of 112 batteries with $d_{1:2}$ are:

Table 4.5: Number of batteries in groups according to availability of measured drift rates

Groups	Drift Rate Availability	Number of batteries
A	Measured $d_{1:2}$ and $d_{2:3}$ exist	18
	Measured $d_{1:3}$ exist	2
B	Measured $d_{1:2}$, no $d_{2:3}$	93
C	No measured drift rate	20
Outliers	Various	4
TOTAL		137

**Figure 4.4:** Histogram of measured drift rates between time points (1) and (2), and between time points (1) and (3).

$$\bar{x}_{1:2} = 0.000291$$

$$s_{1:2} = 0.000014$$

The three batteries with high drift rates are of the order of 17 standard deviations away from the mean. Therefore these three batteries and associated observations were treated as outliers.

The fourth outlier was apparent when comparing $d_{1:2}$ and $d_{2:3}$ for the 19 batteries for which it is possible to calculate both, shown in Figure 4.5. Although there is slight variation of drift rate between the two periods for all batteries, battery 104, towards the top left of the plot, has a much more pronounced difference in drift rates. This battery was also treated as an outlier. Table 4.6 summarizes the batteries classed as outliers.

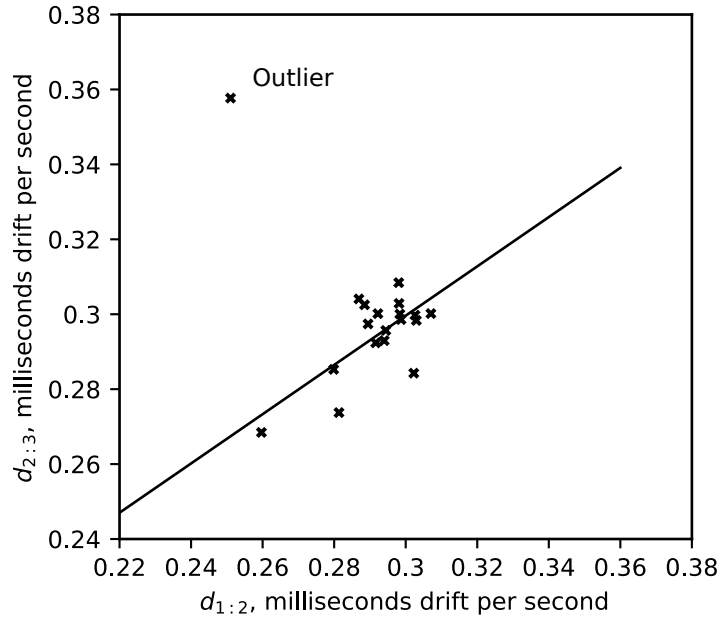


Figure 4.5: Plot of measured drift rates between time points (1) and (2) against measured drift rates between time points (2) and (3). Line of best fit calculated using ordinary least squares, equation (4.18).

Table 4.6: Batteries classified as outliers with reason for classification

Batteries	Reason
2, 24, 137	High measured $d_{1:2}$
104	Pronounced change in drift rates between time periods

4.4.2 Drift Rate Estimation

The method in Figure 4.3 was used on the existing data to infer drift rates for Groups B and C. For the hypothesis test for equal means, the calculation proceeded as follows. Outliers were not included.

$$n_{1:2} = 111$$

$$n_{2:3} = 18$$

$$\bar{x}_{1:2} = 2.91 \times 10^{-4}$$

$$\bar{x}_{2:3} = 2.95 \times 10^{-4}$$

$$s_{1:2}^2 = 1.96 \times 10^{-4}$$

$$s_{2:3}^2 = 1.11 \times 10^{-4}$$

$$t_s = -1.28$$

$$v = 28$$

The critical value for a two-tailed test with these parameters is ± 2.048 . The null hypothesis that the population means $\mu_{1:2}$ and $\mu_{2:3}$ are the same cannot be rejected at the 5% level. Therefore, for batteries in Group C, the drift rate was estimated as the population mean drift rate, μ , which is itself estimated by the sample mean, (equation (4.13)):

$$\bar{x} = 0.000292$$

with standard deviation $s = 0.000014$.

To test whether $d_{1:2}$ and $d_{2:3}$ are unrelated, the regression coefficients β and α were estimated for equation (4.14). The usual process would be to use a sample of the observed data (for example, 80% of it) to generate the model, and then evaluate the model by testing it on the remainder of the data. However, in this case, the number of observations is small ($n = 18$), and if a sample was taken, the result of the model would be influenced by the set of observations selected. Therefore, all the data are used to construct the model.

$$\hat{\beta} = 0.658, \hat{\alpha} = 0.000102$$

Equation (4.18) gives the predictive model for $d_{2:3,i}$ for battery i :

$$\hat{d}_{2:3,i} = 0.658d_{1:2,i} + 0.000102 \quad (4.18)$$

The t-test to investigate the significance of the regression coefficient β of $d_{2:3}$ on $d_{1:2}$ proceeded as follows:

$$n = 18$$

$$t_s = 3.81$$

$$v = 16$$

The critical value for a two-tailed test with these parameters is 2.12. The null hypothesis that $\beta = 0$ can be rejected. Therefore, for Group B batteries, $\hat{d}_{2:3,i}$ was estimated according to equation (4.18) to take account of the correlation.

Also using a t-test, the intercept α was found to be borderline significant, so was included in subsequent calculations.

We show an example of the use of the linear model by estimating $\hat{d}_{2:3,i}$ using equation (4.18) for a battery with a measured $d_{2:3,i}$. For battery 145:

Table 4.7: Summary of drift rate estimates for each group of batteries

Groups	Drift Rate Availability	Drift Rate for Use
A	Measured $d_{1:2}$ and $d_{2:3}$ exist	Use measured $d_{1:2}$ and $d_{2:3}$
	Measured $d_{1:3}$ exist	Use measured $d_{1:3}$
B	Measured $d_{1:2}$, no $d_{2:3}$	Find $\hat{d}_{2:3}$ using equation (4.18)
C	No measured drift rate	Use estimated population mean μ
Outliers	Various	See Section 4.8

Table 4.8: Measured drift rates for batteries classified as outliers

Battery	$d_{1:2}$	$d_{2:3}$
2	0.000535	Use μ
24	0.000541	Use μ
137	0.000558	0.000289

$$\hat{d}_{2:3,145} = 0.000273, \quad d_{2:3,145} = 0.000268$$

The error of the estimate as a proportion of the measured $d_{2:3,145}$ is 1.7%.

Table 4.7 summarizes the drift rates for use in time correction for each group of batteries.

For the three outlier batteries with much higher measured drift rates between time points (1) and (2) than the rest of the sample, it was not possible to know why the drift rates were so high. Possible reasons are that the batteries experienced a different environment to the rest of the population, perhaps undergoing higher or lower temperatures, affecting the speed of the RTCs on the microcontrollers, or that human error in recording time points occurred in the same way for all three. Table 4.8 shows the measured drift rates available for these three batteries. Where measured $d_{1:2}$ and $d_{2:3}$ exist they were used to correct the time readings. Given that for battery 137, $d_{2:3,137}$ drops down to a value within the range of the other observations, for batteries 2 and 24 which do not have measured $d_{2:3}$ observations, the estimated population mean μ was used (equation (4.13)).

Out of the 117 batteries with measured $d_{1:2}$ or $d_{1:3}$ observations, these three outliers comprise 2.6% of the sample. For the remaining 20 batteries without measured $d_{1:2}$ observations in Group C, according to that proportion, it would be expected that 0.5 batteries would have an unusually high $d_{1:2}$. This is acceptably low to ignore the possibility of it occurring for the remaining 20 batteries.

Table 4.9: Time points used for time correction calculations

Location of T_i	Drift Rate	Values for T_a and t_a
Time point (1)–time point (2)	$d_{1:2}$	T_1, t_1
Time point (2)–time point (3)	$d_{2:3}$	T_2, t_2
Time point (1)–time point (3)	$d_{1:3}$	T_1, t_1

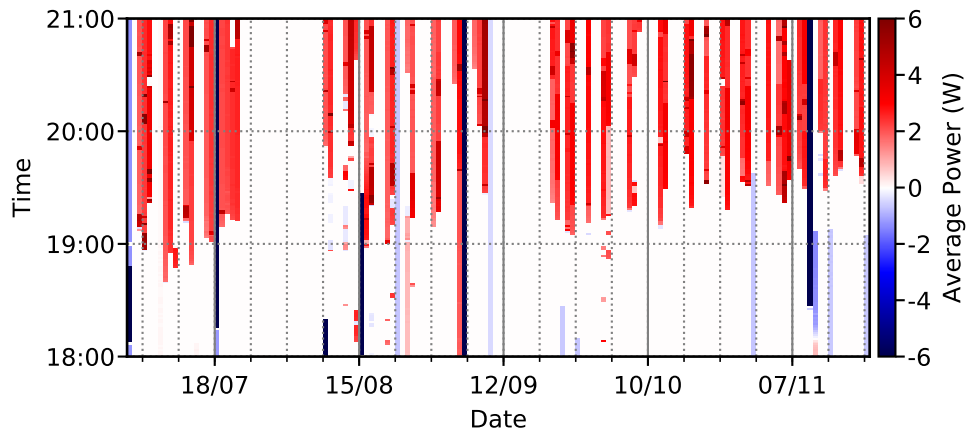
Battery 104 was counted as an outlier due to the pronounced change in drift rate observations between the two time periods. It is not possible to know why the change in drift rate was so pronounced, and so the measured drift rates were used in time correction for this battery.

4.4.3 Time Correction

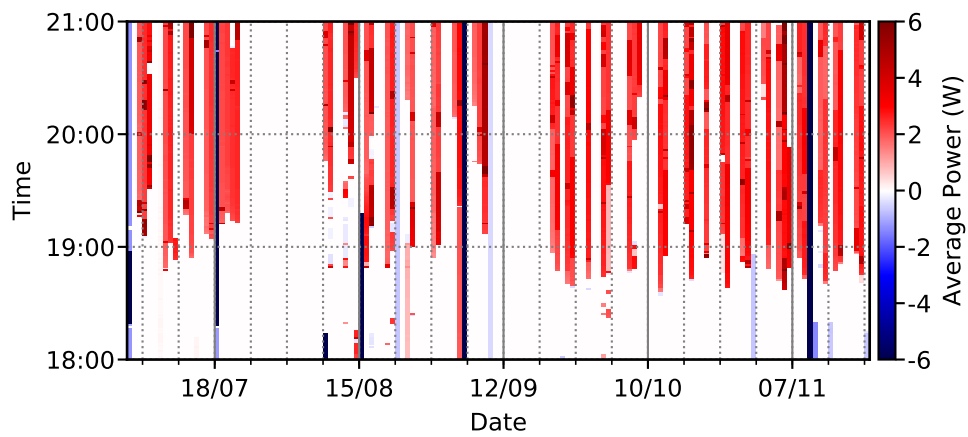
Each battery time point T_i was corrected to t_i using equation (4.6). The equation requires a recorded set of readings of battery time and equivalent actual time, T_a and t_a . Table 4.9 indicates which time points were used as reference times T_a and t_a , depending on when in the trial the battery time reading T_i exists. All data sets were referenced to Eastern Africa Time (EAT, GMT+3) so that patterns displayed can be more intuitively understood. Figure 4.6 shows heat maps for before (a) and after (b) time correction for an indicative battery. The heat maps show electricity data for the whole trial period between 18:00–21:00 to focus on the time electricity starts to be used in the evening, as this best shows the impact of the time correction. It is clear that throughout the trial, for the raw data (Figure 4.6a), the time at which electricity starts to be used in the evening gets increasingly late. After correction (Figure 4.6b), the time that electricity starts being used is more consistent, occurring around the time of sunset.

4.4.4 Accuracy of Time Correction

The accuracy of correcting T_i to t_i for the various cases of correction using different drift rates can be calculated. For measured drift rates, the uncertainty in the drift rates is due to rounding error in the time point recordings, and the uncertainty in t_i can be calculated as uniform error bounds around t_i . For predicted drift rates $\hat{d}_{2:3}$, the uncertainty comes from the model uncertainty in the prediction of $d_{2:3}$, and can be expressed as confidence intervals around the predicted drift rates. This can be propagated through the time correction calculation to observe the uncertainty on t_i caused by these confidence intervals. Similarly, for the mean drift rate, confidence intervals around the mean drift rate can be calculated and the affect on t_i examined.



(a) Before time correction



(b) After time correction

Figure 4.6: Heat maps for an indicative battery showing data a) before and b) after time correction. Heat maps focused between 18:00–21:00 for clarity.

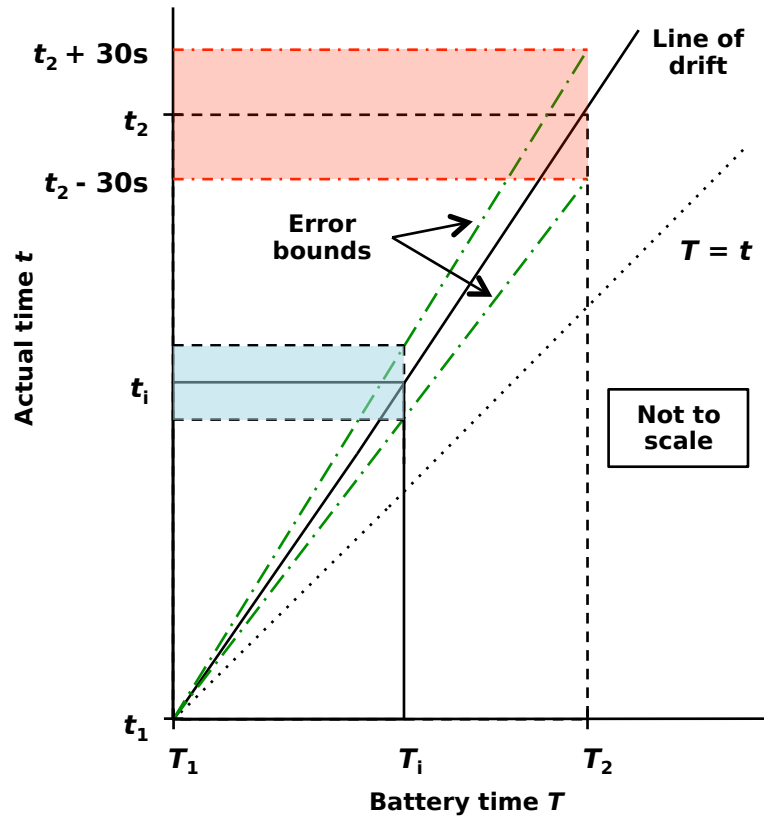
4.4.4.1 Accuracy of Time Correction with Measured Drift Rates (Group A)

The time points recorded throughout the data collection period were rounded to either the nearest second or the nearest minute. This creates uniformly distributed measurement accuracies for each time point with bounds shown in Table 4.10. Given we are not working to a higher degree of accuracy than one second, the rounding errors of ± 0.5 seconds are ignored, so rounding error is relevant for t_1 and t_2 only.

Figure 4.7 is a schematic of the relationship between the actual time and the battery time, showing a visualization of the error bounds involved in time correction for battery times between T_1 and T_2 using measured drift rate $d_{1,2}$. The line $T = t$ is shown for reference, to contrast with the line of drift. The measurement error bounds on t_2 , shown with dashed red

Table 4.10: Rounding error bounds on time point readings

Time Point	Rounding Error
T_1, t_1	
T_2	± 0.5 seconds
T_3	
t_2	
t_3	± 30 seconds

**Figure 4.7:** Error bounds in battery time correction between time points (1) and (2) using measured drift rate $d_{1:2}$.

lines and shading, mean that there is an error bound associated with the $d_{1:2}$ estimate, so the actual relationship between T and t lies within the error bound lines in green. Therefore, when correcting battery time T_i , t_i is found with an associated error bound, shown in shaded blue. For the case of time correction of battery time readings $T_1 < T_i < T_2$ using measured $d_{1:2}$ values, the error bounds on t_i get larger as T_i increases, to a maximum of ± 30 seconds at $T_i = T_2$.

Where measured $d_{1:3}$ is used for battery time readings $T_1 < T_i < T_3$, the schematic in Figure 4.7 applies but over the larger time span between time points (1) and (3). The error bounds on t_i will increase as T_i increases, to a maximum of ± 30 seconds at $T_i = T_3$, which is the accuracy of the t_3 reading.

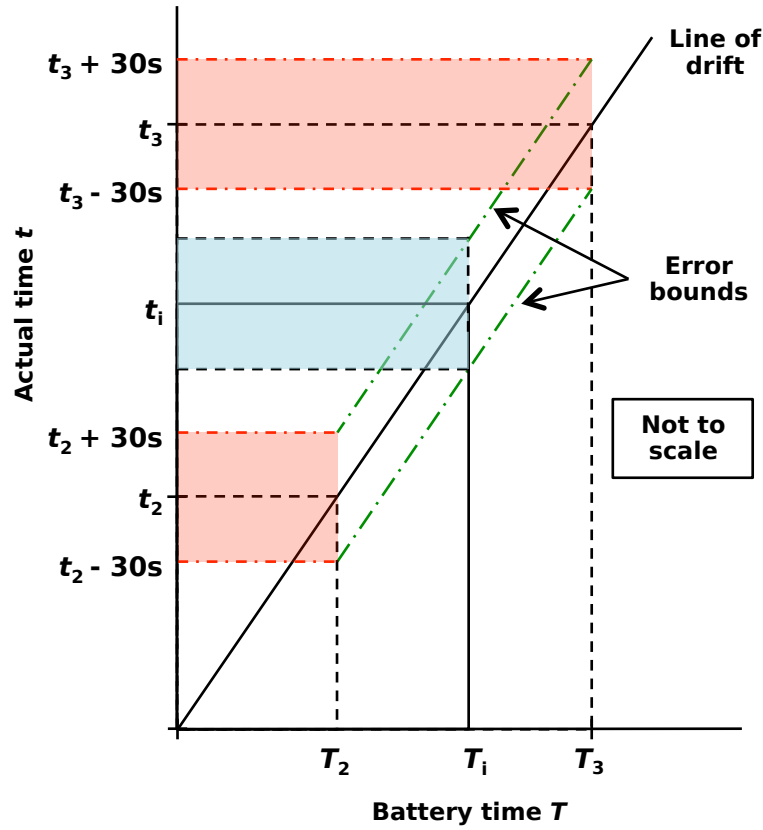


Figure 4.8: Error bounds in battery time correction between time points (2) and (3) using measured drift rate $d_{2,3}$.

Figure 4.8 shows a similar schematic for time correction of battery time readings $T_2 < T_i < T_3$ when measured $d_{2,3}$ values are used. This time both t_2 and t_3 have error bounds of ± 30 seconds, shown in red, and so the resultant error bounds on t_i are constant for any T_i in this period.

Thus time correction using measured drift rates result in very accurate actual time estimates, with a maximum error bound on the corrected time of ± 30 seconds.

4.4.4.2 Accuracy of Time Correction with Predicted Drift Rates (Group B)

The predictive model to estimate $d_{2,3,i}$ is given by equation (4.18). The observations of $d_{1,2}$ are approximate observations of the true value of the drift rate for the first period between time points (1) and (2). Performing the regression analysis of $d_{2,3}$ on $d_{1,2}$ using these observations therefore causes regression dilution (also known as attenuation bias). This is where the estimate of the regression slope $\hat{\beta}$ is biased towards 0 relative to the true value of β that corresponds to the regression of $d_{2,3}$ on the true values of $d_{1,2}$. In this case the bias of the estimate $\hat{\beta}$ is not a problem because the model is being used for predictive purposes,

estimating the value of $d_{2:3}$ that corresponds to the observation of the independent variable $d_{1:2}$ that is known to be approximate. Correction of the value of $\hat{\beta}$ is therefore not necessary.

The prediction of $d_{2:3}$ is uncertain because of the model error term – the fact that equation (4.18) does not perfectly explain $d_{2:3}$. The standard deviation of the errors in the regression analysis can be used to construct confidence intervals for the predictions of $d_{2:3}$. However, the estimates $\hat{\beta}$ and $\hat{\alpha}$ also have standard errors associated with them. They are estimates of the true β and α using the 18 observations of $d_{1:2}$ and $d_{2:3}$. The uncertainty in $\hat{\beta}$ and $\hat{\alpha}$ (parameter estimation uncertainty) can be included in the estimated standard error for predicting $d_{2:3}$ from $d_{1:2}$ by calculating $\hat{\phi}$ as shown below. $\hat{\phi}$ is dependent on $d_{1:2,i}$ and so varies slightly for each battery.

$$\hat{\phi} = \sqrt{[1 \quad d_{1:2,i}] \begin{bmatrix} \text{var}(\hat{\alpha}) & \text{cov}(\hat{\alpha}, \hat{\beta}) \\ \text{cov}(\hat{\alpha}, \hat{\beta}) & \text{var}(\hat{\beta}) \end{bmatrix} \begin{bmatrix} 1 \\ d_{1:2,i} \end{bmatrix} + \text{var}(\hat{\epsilon})} \quad (4.19)$$

Assuming the model error term is normally distributed, 95% confidence intervals for each drift rate prediction can be written as:

$$P(\hat{d}_{2:3,i} - 1.96\hat{\phi} < d_{2:3,i} < \hat{d}_{2:3,i} + 1.96\hat{\phi}) = 0.95$$

The $\hat{d}_{2:3}$ values will be used to correct battery time readings $T_2 < T_i < T_3$. Figure 4.9 shows a schematic showing the 95% confidence interval bounds for time correction in this case. The uncertainty in the predicted drift rates mean there is a range of possible gradients for the line describing the relationship between the battery time T and actual time t , the extremes of which are positioned on the maximum values of the rounding error of t_2 (shown by the red dashed lines and shading). The uncertainty in t_i therefore increases as T_i increases, shown by the widening of the green dashed lines, and the increased width of the error bounds for t_i (blue) compared to t_2 (red). To show the worst-case uncertainty in t_i estimates, the 95% confidence intervals on t_i can be found when $T_i = T_3$. Figure 4.10 shows these confidence intervals for the 93 batteries that use equation (4.18) to predict $d_{2:3}$, with all worst-case confidence intervals on t_i being under 5 minutes. When $T_i < T_3$ the confidence intervals will be smaller, ranging linearly from ± 30 seconds at $T_i = T_2$ to the upper values shown in Figure 4.10.

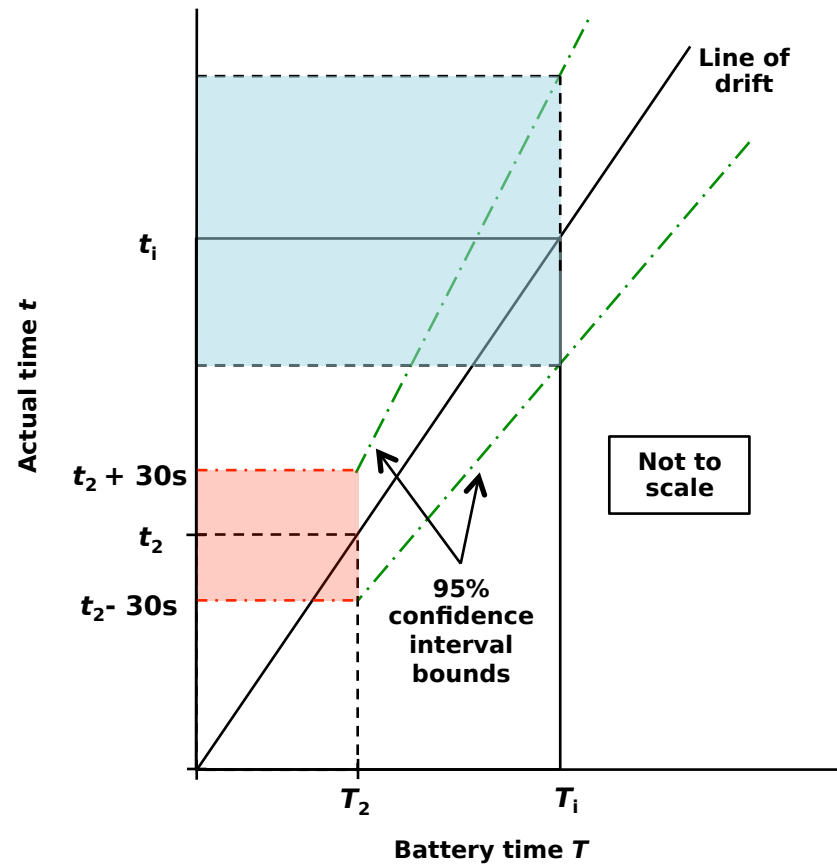


Figure 4.9: 95% confidence intervals on battery time correction using predicted drift rate $\hat{d}_{2,3}$, for battery times between time points (2) and (3).

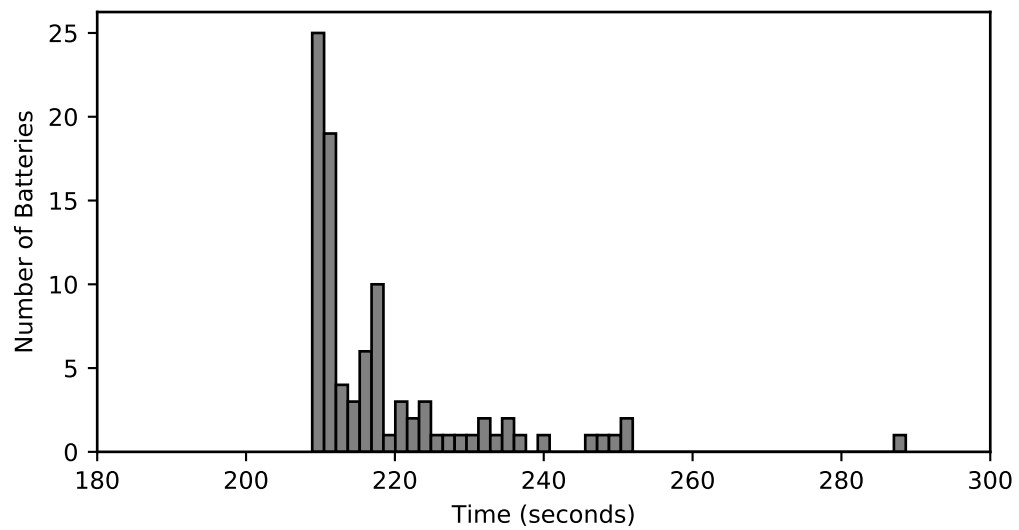


Figure 4.10: Size of 95% confidence intervals in corrected time t_i using predicted drift rates $\hat{d}_{2,3}$ at time point (3).

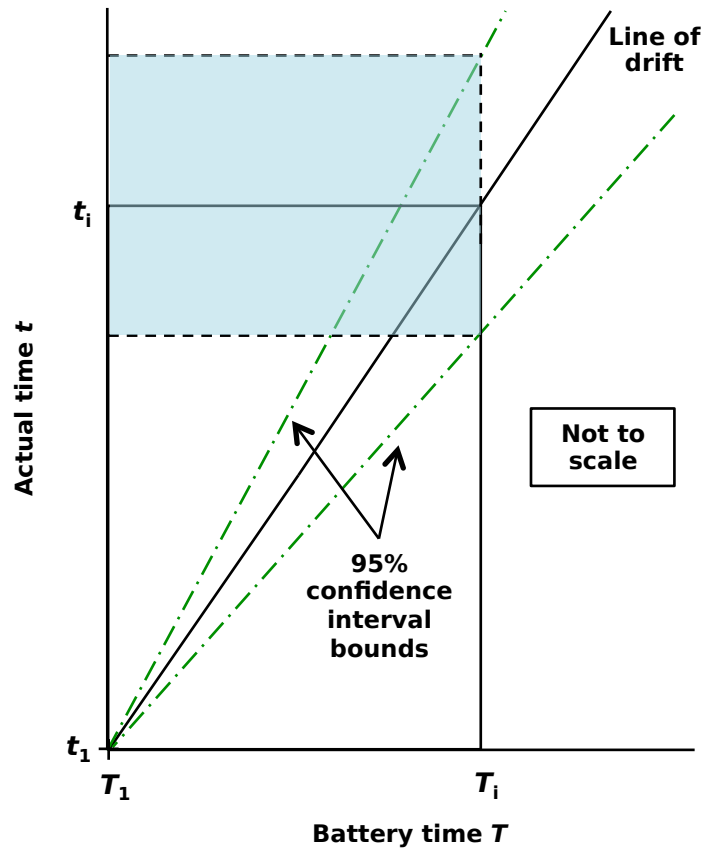


Figure 4.11: 95% confidence intervals on battery time correction using mean drift rate.

4.4.4.3 Accuracy of Time Correction with Mean Drift Rate (Group C)

For batteries in Group C, for which there are no measured drift rates, the estimate of the population mean of all drift rate observations is used in the time correction. Confidence intervals can be constructed around the mean drift rate, similarly to the case of predicted $d_{2:3}$ values above.

Assuming that the error in the mean drift rate is normally distributed, 95% confidence intervals for the mean drift rate are defined by the standard formula:

$$P\left(\bar{x} - 1.96 \frac{s}{\sqrt{n_{1:2} + n_{2:3}}} < \mu < \bar{x} + 1.96 \frac{s}{\sqrt{n_{1:2} + n_{2:3}}}\right) = 0.95$$

Figure 4.11 shows a schematic showing how the confidence intervals about the mean drift rate affect the time correction of T_i to t_i . Similarly to the case for predicted $d_{2:3}$ values, there is a range for possible gradients for the line that describes how battery time relates to actual time. However, in this case, there is no measurement error on the starting time point (t_1 and T_1), so the green lines start together. The 95% confidence intervals on the corrected time t_i therefore increase with increasing time passed since the first time point.

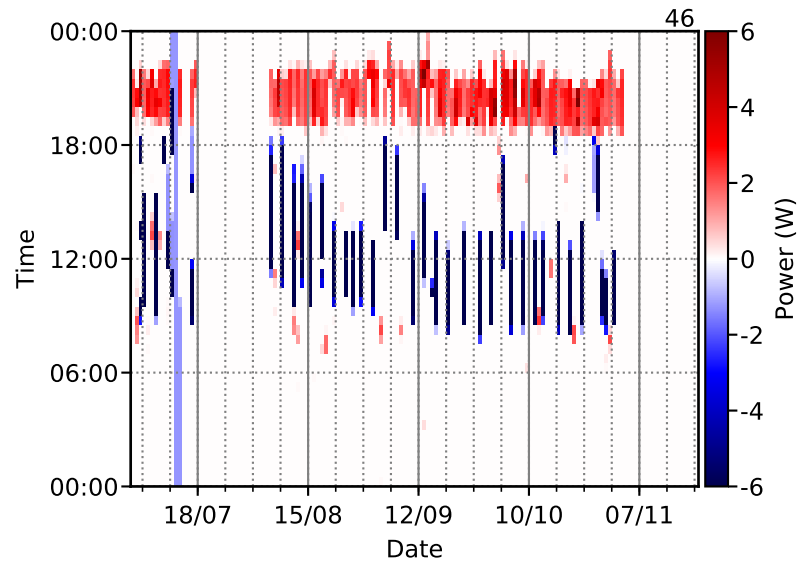


Figure 4.12: Heat map of processed electricity data for one of the households (household 46).

In this case, there is only one drift rate applied to all batteries in Group C (the estimate of the population mean). The effect of the uncertainty in the mean drift rate can be calculated on t_i for a worst-case situation. The worst-case of uncertainty in t_i occurs at the longest period of time elapsed from the first time point, which is at time point (3). We use the average time span between time points (1) and (3) for batteries with both measurements, which is 261 days 13:54:21. Over this period, the uncertainty in t_i , calculated using 95% confidence intervals on the mean drift rate, is ± 53 seconds.

4.4.5 Final Result: Household Electricity Profiles

Following the time correction by the methodology demonstrated above, as well as the 12 additional processing steps described in Table 4.2, the resulting data set consists of average power each half-hour for the 12V port of the household batteries attributable to 51 households. This is available on Mendeley Data at the doi noted in the Preface of this thesis. Figure 4.12 is a heat map of the electricity data for one of the households; the remainder are in Appendix C.

Visual inspection of the data on the USB port showed that the mobile phone charging functionality was seldom used, as households prioritized using the battery for light. Given the electricity use on the USB port was so low, and the fact the data readings included power to the microcontroller, the USB data set was not cleaned and transformed into electricity profiles.

Step 12 in Table 4.2 involved converting the electricity data to profiles with half-hour periods. The electricity data are also released in this form. Half-hour intervals were chosen for

several reasons. From a technical perspective, 30 minute resolution is sufficient for systems level design that involves sizing generation, storage, and distributions components (i.e.: not inclusive of power electronics). In terms of prior convention, it is also a standard monitoring period for smart meter data. In addition, it aggregates electricity use enough such that detailed behaviour cannot be inferred, in contrast to data at higher time resolutions (e.g.: 1 minute), which might be used in a way that is invasive of customers' privacy [87], [141]. However, it does still enable occupancy patterns to be inferred, so the data are fully anonymized to make it impossible to link individual households from either community to the electricity data.

4.5 Discussion

4.5.1 Quality of Electricity Profiles

Analysis of the uncertainty around the measured and estimated drift rates, and the resulting accuracy of time corrections, shows that the most uncertain time corrections occur when using predicted drift rate values to correct data between time points (2) and (3). However, even then, the 95% confidence intervals around the predicted drift rates lead to uncertainty in corrected t_i of less than ± 5 minutes. These uncertainties are small enough to ignore in further analysis of the data.

For the case study, the difference between assuming $d_{2:3} = d_{1:2}$ compared to using the predictive model (equation (4.18)) is in fact small – a difference in the time correction of 10 minutes over a year. However, it could be important for monitoring trials of longer duration, or in cases where the drift rate is more severe. The aim of this research contribution is to provide a methodology that is more generally applicable, rather than just applicable to the case study. In addition, the linear regression analysis and hypothesis test procedure is standard and easily applied in statistical software packages, such as using the module `numpy` in Python.

The other steps in Table 4.2 were carried out to retain as much of the raw data as possible, while retaining data integrity. The issue of the reset of the battery RTCs to 1970 (step 3) caused a lot of data to be discarded because it was not possible to calibrate it back to the correct times. It was only possible to correct the portion of data that had this issue for 9 out of 49 batteries. Although it would have been possible to guess more of the calibrations, it was decided that it was preferable to proceed only when calibrations were almost certainly correct, rather than make calibrations that reduced the accuracy of the data set and would

have resulted in large uncertainties that extended across multiple days. The issue with the 1970's resets show very clearly the impact of not having real-time time synchronization functionality in the sensor system. For a case such as this, where frequent repairs to the technology carrying the sensors caused the RTC resets, time synchronization is more crucial than if the technology is reliable throughout the trial period.

Another source of difficulty, and where careful judgment was sometimes required, was in assigning batteries to particular households. Step 7 contributed to this, where SD cards that had been swapped between batteries required assignment to the correct battery; step 8 was also important, where batteries were assigned to particular users; step 11 was the final stage, where multiple user profiles were combined to give household specific profiles. At each stage this was an arduous process using records kept by hand by the technicians working in the hubs, which were in the most part a complete record of battery assignments. However, occasionally judgments were required where the evidence was not conclusive, or where contradictions occurred. These were made according to the best knowledge available. Combining user profiles into single household profiles (necessary when a household used two batteries at once) was carried out under conservative assumptions, so although some data were discarded, the data that were kept were known to represent all the electricity use through the battery-charging system for each household.

Step 9 addressed logging disruptions that occurred during charging, filling in typical charging profiles where appropriate. Step 10 set power and duration of use limits, eliminating days where the battery had been used for under 10 minutes, had not been used to power an actual load, or had just been left at the hub having been fully charged. Both of these steps ensured that the data set was representative of when significant energy was in use (whether charging or powering loads), and did not include days where the battery system was not actually used - perhaps being switched on to check the charge level or waiting for collection from the hub. This was important so that in subsequent analysis, actual use days were not diluted through the inclusion of days that were not representative of use of the system. However, both of these steps also mean that some of the practices of the end users are no longer visible in the data, for example, the time batteries were collected from the charging hub. It was decided to prioritize the electricity data, rather than the information that could be drawn about practices, though obviously the latter is also useful when it comes to understanding more about how the system is used.

4.5.2 Use of the Data in Further Work

The electricity profiles constructed in this chapter will be utilized throughout the remaining chapters of this thesis. The following text discusses known weaknesses of the data set, and caveats to the use of the data for different contexts and electricity access systems. This discussion is also relevant for those who may want to utilize this data set in their own design or research.

Firstly, the data set is small in size compared to the thousands of households that tend to make up smart meter electricity data sets. It consists of electricity data from 51 households, with measurement periods between 5 days and 5 months. The reasons for this include: the inherent small scale of community projects; an unanticipated reduction in the size of the roll-out, due to both technical issues with the household batteries and insolvency of the NGO project partner midway through the project; and the limited resources available to this research project, which did not enable development of more systems. A further caveat is that the participants were self-selecting, and as such they are not a random sample from the communities – they are early adopters of a new electricity system. However, survey data show that they represent a range of households in the villages in terms of demographic characteristics, and it is likely that lighting needs of early-adopters are representative of needs throughout the village. In addition, in [142] it has been shown that when aggregating load profiles of Global North residential customers, diversity and variance in the load profiles fall rapidly with increasing group size, levelling out from a group size of approximately 10. Therefore, the data set is of large enough to contain sufficient diversity of use, and throughout this thesis it will be shown that it is sufficient to pick up typical use patterns within and across the households. It is a small but high quality data set, which represents the range of electricity use across and within the two villages.

Secondly, the data set aggregates electricity data from two rural communities in Nakuru County, Kenya. Chapters 5 and 7 investigate demographic links to electricity use, which implicitly assumes that the links are the same for both communities. The option existed to keep the data sets separate by community, but the sample sizes were deemed too small: 23 households and 28 households respectively. In addition, there are many similarities between the communities in terms of their location in the rural area around Nakuru, their

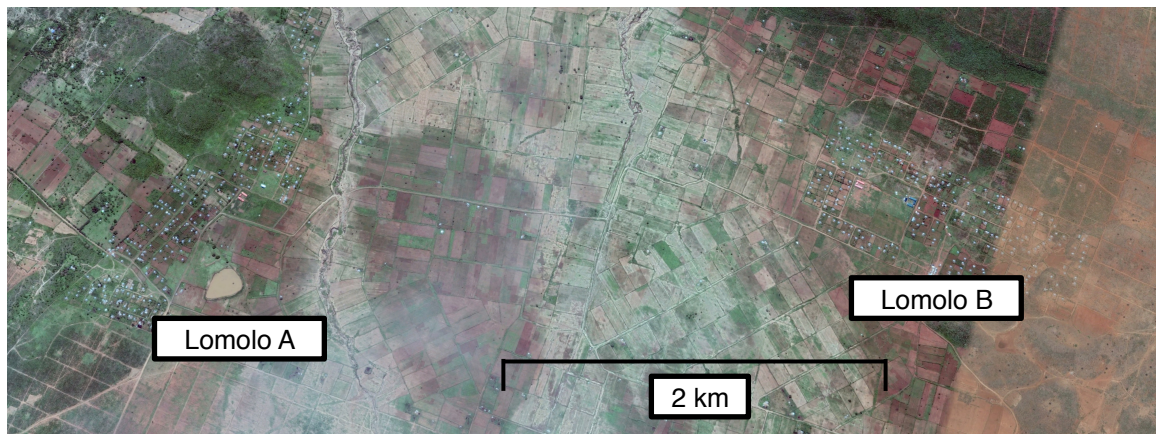


Figure 4.13: Google earth image showing Lomolo A and Lomolo B, showing approximate scaling. Lomolo A is a nearby and socioeconomically and culturally similar village to Lomolo B.

livelihoods, and other factors that contribute to similar energy cultures, so this is not an unreasonable assumption to make.

In the following chapters, the data are utilized to demonstrate methodologies to estimate electricity demand for electricity systems for new communities. This raises the question as to the external validity of the data, in terms of the transferability of it outside of Lomolo B and Echariria, and generalizability to communities in different geographical locations. Energy behaviour depends on many factors, internal and external to the household, as characterized by the Energy Cultures Framework, for example (see Section 2.3.2, Chapter 2). Whether or not the data are applicable to a particular context must be carefully judged; if the energy cultures of Lomolo B and Echariria are similar to the new community, the data are applicable. An example of where the electricity data set presented here is evidently applicable elsewhere is for Lomolo A, a neighbouring village to Lomolo B (Figure 4.13). Located 3 km away in the same sisal plantation, residents from both Lomolo A and B originated in the community in the Mau forest before resettlement. Residents are culturally and socioeconomically similar, and local context, including infrastructure and supply chains, are the same.

The transferability of the data to other electricity access systems has several considerations. The electricity data set presented here is for a lighting service constrained to three 3W lights, so does not reflect the use of other services powered by higher energy and higher power micro-grids. However, lighting is a basic service that is included in most electricity access systems, and is the dominant service for SHSs and other pico-solar products. Therefore, the data can be used to model demand for lighting, and even where systems are designed to provide other services, it can reduce some of the unknown nature of demand for higher power,

higher electricity systems. (Use of the data set to estimate demand for higher electricity access tiers according to the MTF is explored in Chapter 7.)

Another caveat is that the architecture of the battery-charging system may place constraints on electricity use that would not be present for SHSs or micro-grids. For battery-charging systems, there is an additional cost to electricity use in the time and effort taken to charge the battery. This cost may mean households use less electricity, in order to charge the battery less often. If this is the case, when that constraint is lifted, as it is for a SHS or a micro-grid, households may use more lighting electricity, such that the data set presented here is an underestimate for these systems.

A way to examine whether or not this might be the case, looking at data from this study alone, is to test whether there is a correlation between electricity used, and distance of the household to the charging hub. Distance of the household to the hub represents the time and effort cost of charging, though it is recognized that in reality this is likely to be multi-factorial, and may also include factors such as how able-bodied household occupants are, and the number of household occupants. Average daily electricity consumption and distance can be extracted from the case study electricity data, and user surveys obtained self-reported time taken to travel to the hub for 38 households. If households economize on their electricity use, and charge less frequently due to distance from the hub, we might expect a negative correlation between distance and daily electricity use. Figure 4.14 shows the average daily electricity consumption against the travel time to the hub for each household, along with a best fit line calculated using ordinary least squares, and a p-value of a hypothesis test for the significance of the correlation coefficient. The null hypothesis for no correlation can be rejected to a significance of 5%, but for a more conservative significance of 1%, it cannot be rejected. This suggests electricity use does depend on distance to the hub, at conventional 5% significance levels.

The issue is further complicated by financial costs of electricity systems. For micro-grids, electricity is usually paid for in proportion to the amount used. Typically, for battery-charging systems and for SHSs, a subscription fee is paid periodically and is independent of actual electricity use (for a given battery capacity or system specification, which would of course influence the subscription cost). Therefore, for a micro-grid (and similarly if a battery-charging-system were operated on a pay-per-charge model), electricity use may be constrained by financial cost, rather than the time and effort cost that is dominant for a

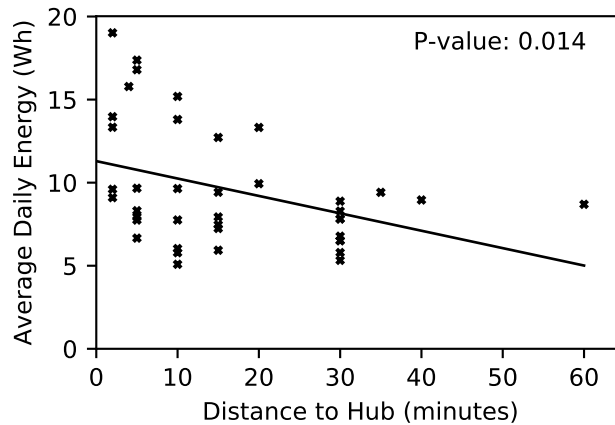


Figure 4.14: Plot of average daily electricity consumption per household against distance to the hub in minutes, with line of best fit calculated with ordinary least squares regression. P-value for a hypothesis test for the significance of the correlation coefficient.

battery-charging system on a subscription cost model once the subscription fee has been paid. It is challenging to compare financial cost of electricity for a micro-grid with time and effort cost of a battery-charging system.

The case study reflects the situation of a battery-charging system with a periodic subscription payment model. This is despite there being no financial cost to the end user for the battery-charging system during the monitoring period, because the planned periodic subscription fee was waived due to the technical issues. However, we argue electricity use would be unaffected by this, in comparison to if there had been a subscription fee as was planned, and as was implemented after the monitoring period with the deployment of the larger lead-acid batteries. This is because a periodic subscription fee is independent of the actual daily electricity used, and so serves to dictate whether or not a customer chooses to be part of the system at all, but not their behaviour once they are enrolled. The end users in the case study signed up to the system expecting a subscription fee, so are reflective of actual customers who would have paid.

To summarize, the characteristics of the physical architecture of different electricity access systems, and the financial mechanisms associated with them, may affect electricity use. We have shown this to be true for the distance of the household from the charging point, for example. This of course might limit the relevance of the results for one system for another system. A direct comparison between lighting electricity data from a battery-charging hub and an alternative system (in which electricity is paid for per unit used) from comparable

contexts would help to shed light on this. In the meantime, this data set should be used cognizant of the caveats to transferring the data to new electricity systems.

4.6 Summary

The research question addressed in this chapter was: How can the accuracy of time series electricity data collected from multiple sensor nodes be improved when real-time time synchronization is not possible, and there is limited time calibration data available?

A method to estimate drift rates for sensors without measured data has been outlined, using two statistical tests in series to indicate what the best estimates for unknown drift rates are. The time correction has been demonstrated for lighting electricity data collected from household batteries in the solar nano-grid case study, and the data set was corrected to within ± 5 minutes of the actual time (discounting outlier batteries).

This technique was used alongside 12 further processing steps to construct 51 household lighting electricity profiles from the raw electricity data from the batteries. These will be used in subsequent analysis to demonstrate the methodologies throughout this thesis, and are available to support further research and development.

How can electricity demand data and demographic data be used to estimate daily electricity consumption?

5

Estimation of Daily Electricity Consumption for Electricity Access Systems

Contents

5.1	Introduction	115
5.2	Method	117
5.2.1	Required Data	117
5.2.2	Data Analysis	118
5.2.3	Demand Estimation	120
5.2.4	Inferring Demand for Sizing Battery-Charging Systems	120
5.3	Case Study	123
5.3.1	Demographic Data Acquisition	123
5.3.2	Analysis: Demographic Data and Electricity use	125
5.3.3	Demand Estimation: Community X	128
5.4	Discussion	129
5.4.1	Discussion and Interpretation of Case Study Results	129
5.4.2	Sizing Key Parameters for Battery-Charging Systems	132
5.5	Summary	138

5.1 Introduction

The previous chapter outlined a methodology to correct the timestamps of time series electricity data gathered by distributed sensors for which real-time time synchronization was not possible. This method was used, alongside 12 other steps, to process time series electricity

data gathered from the case study, resulting in electricity profiles for 51 households. This chapter seeks to answer the question: ‘how can time series electricity data and demographic data be used to estimate daily electricity consumption?’, and the data set from the previous chapter is used to demonstrate the methodology. The two steps to this process, 1) analysis of existing electricity data to understand electricity use, and 2) linking electricity use to demographic data so that demand can be estimated for new communities, are both covered in this chapter.

The gap analysis in Chapter 2, Section 2.2, identified that existing work in estimating aggregate metrics of electricity demand, using econometric models, does not account for the variation in customer electricity use. Therefore, the approach taken here retains variation in electricity use for each household by analyzing each household’s electricity use as a discrete probability density function (PDF) of daily electricity consumption. We explore linking these PDFs to single demographic variables, to give PDFs of daily electricity consumption for different types of households. From these, given demographic data for a new community, the PDF of daily electricity consumption for the new community can be estimated. The methodology for this process is outlined in Section 5.2, and demonstrated for the case study electricity and demographic data in Section 5.3. The results from this case study are useful to inform new system design where there is a lack of available data, and support further research on electricity access systems, and are discussed in the first half of the Discussion (Section 5.4.1).

As discussed in Chapter 1, daily electricity consumption is an important demand metric for battery-charging systems. This is because the point of electricity delivery (in the household) is physically separated from the point of generation (at the hub) by the household battery, which is moved between the two. The daily electricity consumption on the household level is important for sizing both the household battery, and the required generation capacity at the hub for battery-charging systems. For example, the size of the household battery depends on the daily electricity consumption in a household, alongside the number of days a battery should last between charges and a depth of discharge factor. The electricity required daily to charge the batteries at the hub depends on the number of batteries present at the hub, and their size. Therefore, in this chapter, once the PDF of daily electricity consumption has been constructed for a group of households, we infer a single value of daily household electricity consumption from the PDF, from which the household battery can be sized. This is described further, along

with the means of inference of the design point, in the Methodology (Section 5.2) and shown for the case study (Section 5.3). In the second half of the Discussion (Section 5.4.2), we then discuss the factors involved in translating this single value of daily electricity consumption, and the PDF of daily electricity consumption, into actual sizing of the household battery and the hub generation capacity for battery-charging hubs. It is not within the scope of this thesis to go beyond this to a detailed design of the sizing of the battery-charging hub components.

5.2 Method

The proposed method consists of two parts:

1. Analysis of electricity and demographic data for an existing battery-charging hub, to identify commonalities between demographic characteristics and electricity use.
2. Demand estimation of daily end-use electricity requirements for a new community, from easily measured demographic data.

5.2.1 Required Data

Electricity data and demographic data are required that are attributable to households using a battery-charging electricity access system. The electricity data consist of a set of daily electricity consumption values for each household. The demographic data are gathered through means of a survey, covering household characteristics such as the number of occupants, and associated ages, genders, and occupations.

From the survey data, household attributes are defined that are single measurements of some aspect of the household, and that can be tested against corresponding electricity data to see if there is a link. Example attributes are: number of occupants per household, average age of household occupant, percentage female occupants. To choose an attribute to carry forward in the analysis, the correlation between average daily electricity consumption and each attribute can be investigated. An attribute showing the strongest correlation with average daily electricity consumption, observed via the correlation coefficient, is selected. It is also interesting to test for correlations between attributes, because the presence of a correlation provides more information about the characteristics of the households in the analysis, which may offer insight into the meaning of the results.

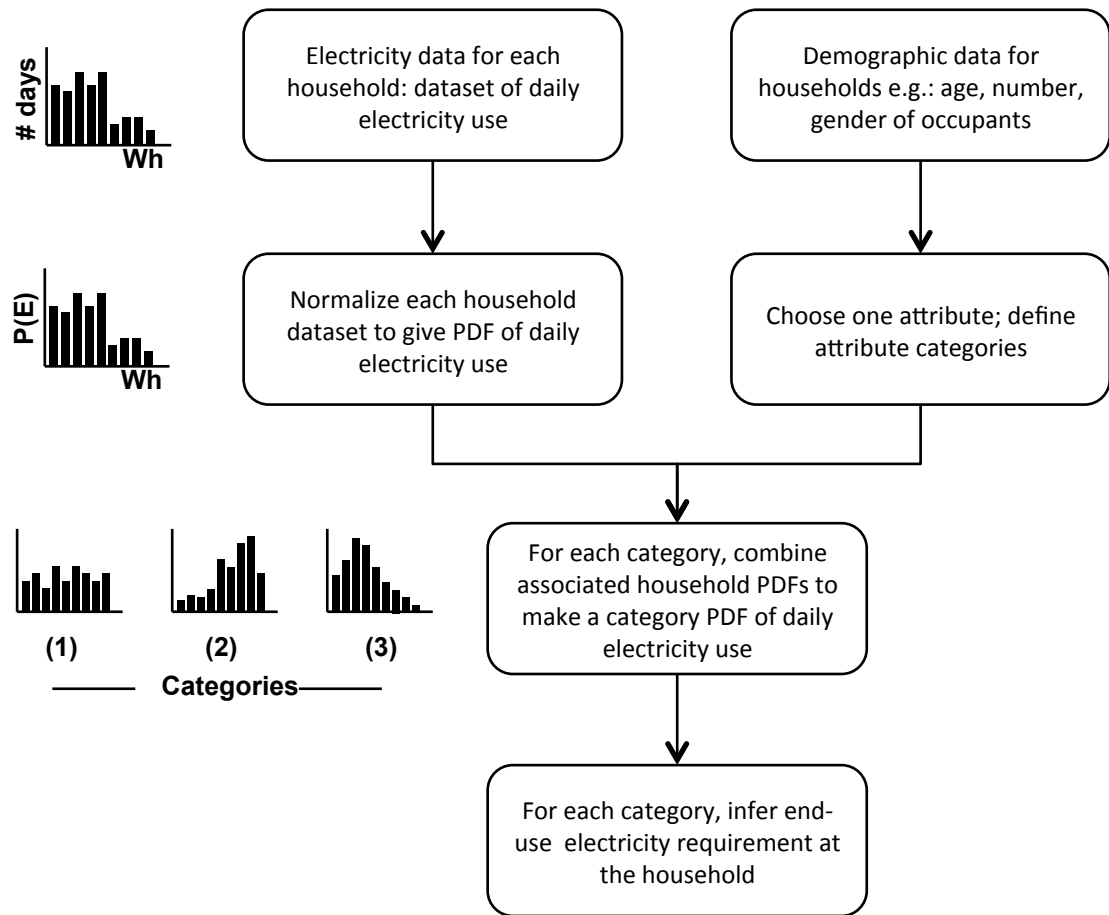


Figure 5.1: Method to analyze electricity and demographic data to give discrete probability density functions (PDFs) of daily electricity consumption for different categories of households.

In preparation for the analysis, the attribute must be split into categories, such that each household is assigned to one category per attribute. A small number of categories is desirable to make results easy to visualize; however, enough categories must be chosen to capture the diversity in the households adequately.

5.2.2 Data Analysis

Both the analysis step and the demand estimation tool involve estimated discrete probability density functions (PDFs) of daily electricity consumption per household, applicable to households, collections of households, or new communities. The PDFs show the probability of a range of magnitudes of daily electricity consumption, and are represented by $P(E)$, where E is energy (Wh). Figure 5.1 shows the method to analyze the household electricity data and find estimated discrete PDFs of daily electricity consumption associated with the categories of a particular household attribute.

The first step calculates PDFs for each household by starting with histograms, showing the number of days for which daily electricity consumption was a certain magnitude. The histograms are constructed by bucketing energy values between whole numbers of watt-hours, inclusive of the lower bound and exclusive of the upper bound. The numerical energy value used to describe each bucket is the midpoint of each interval, which has a fractional digit of 0.5. To convert the histograms into PDFs, the frequency counts are normalized by the number of observations (days of electricity use for that specific household), n_d .

PDFs that represent the distribution of daily electricity consumption for different household categories are calculated by taking an average of the household PDFs that are assigned to that category. The definition of the end-use electricity requirement for each category (used to size a household battery), is described in Section 5.2.4.

For each PDF, it is possible to calculate a cumulative distribution function (CDF). CDFs show the probability that, for a particular daily electricity consumption, less than or equal to that energy magnitude was used on a day. For simplicity, the CDFs corresponding to each PDF are not shown on Figure 5.1.

If the daily electricity consumption is influenced by the value of the household attribute, the category PDFs will have different shapes, showing different distributions of daily electricity consumption between categories. This can be observed by plotting the category PDFs. In addition, to assess similarity more formally, a two-sample Kolmogorov-Smirnov (KS) test [143] can be used. A two-sample KS test is a non-parametric test, which means that it does not assume an underlying distribution. It seeks to determine if two sample distributions differ significantly, through a test statistic (described formally below) based on the maximum vertical distance between the two CDFs of the sample distributions. It can only be used on two distributions at a time, so pair-wise comparisons for all categories are carried out.

The null hypothesis, H_0 , is that the two sample distributions are drawn from the same reference distribution. For two observed CDFs, $F_1(x)$ with number of observations n , and $F_2(x)$ with number of observations m , the test statistic is:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| \quad (5.1)$$

This is compared to the critical value for a significance α :

$$D_{n,m,\alpha} = c(\alpha) \sqrt{\frac{n+m}{nm}} \quad (5.2)$$

where the value of $c(\alpha)$ is given by:

$$c(\alpha) = \sqrt{-\frac{1}{2} \ln \alpha} \quad (5.3)$$

If $D_{n,m} > D_{n,m,\alpha}$, the null hypothesis is rejected, indicating the two sample distributions are assumed to be drawn from different reference distributions. Otherwise, the null hypothesis is accepted. This means that the PDFs are not significantly different, according to the definition of similarity used by the KS test - the maximum vertical different between the two CDFs.

5.2.3 Demand Estimation

The second part of the method is a demand estimation tool to estimate the electricity requirement for new communities, which are expected to have similar electricity behaviour to the existing community used as the learning data set. Figure 5.2 shows the sequence of steps, starting with the acquisition of demographic data for fictional example Community X. PDFs for household attribute categories are combined in proportion to the representation of each category of households in Community X. The resulting community PDF estimates the distribution of daily household electricity consumption in the community.

5.2.4 Inferring Demand for Sizing Battery-Charging Systems

The results of the analysis step of the electricity data (Figure 5.1) are PDFs and CDFs of daily electricity consumption for different types of households, showing the range of magnitudes of electricity used daily within and across the households. The result of the demand estimation step (Figure 5.2), the PDF and CDF for Community X, is an estimate of daily electricity consumption characteristics for that community. In order to observe the effect of different electricity consumption characteristics on the capacity requirement of a household battery for a battery-charging system, we infer from the shape of the PDF/CDF a single value of daily household electricity consumption. This is called the ‘household end-use electricity requirement’, E_x . E_x is a single electricity consumption value that can be used as a benchmark for household battery design.

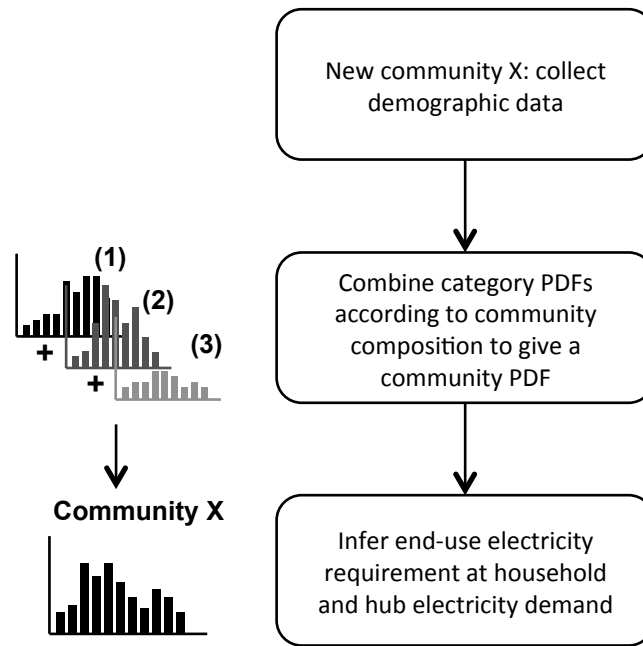


Figure 5.2: Demand estimation tool to estimate discrete probability density function (PDF) of daily electricity consumption for a new Community X.

The y-axis of the CDF shows the percentage of measured or estimated demand that is less than or equal to the magnitude of a given daily electricity consumption value. For a particular E_x , this can also be interpreted as the probability that demand is met at a particular household/group of households on a given day, if up to E_x electricity is accessible for each household (assuming the PDF/CDF describes the electricity use characteristics of that household or group). In analysis by Louie et al. [39], the ‘reliability’ of a system is defined as ‘the ability of a system to provide the demanded load each [relevant time interval]- considering only energy balance constraints’. For ease of explanation, we adopt this definition here for the percentage of demand met, or probability that demand is met, as described above. In this case, the definition is relevant for the ability of the household battery (in terms of capacity) to provide the demanded electricity each day shown by the daily electricity consumption PDFs/CDFs.

The reliability is a variable that can be set in order to extract a value for E_x . For example, by deciding that 100% reliability¹ is necessary, $E_x = \max(E)$. Deciding to allow less than

¹When referring to the PDF/CDF of estimated demand for Community X, in reality it is not possible to design a system that will meet demand 100% of the time. This is due to intermittent generation, the potential for consumption that is higher than predicted, and load growth. In this case, it would be correct to acknowledge this by using terminology such as ‘99.9% reliable’. However, given this explanation is also relevant for the measured PDFs derived from the analysis step (Figure 5.1), for which it is possible to say, retrospectively, that a certain E_x would have met demand 100% of the time, we continue to use ‘100% reliable’ for simplicity. We do so referring to the maximum daily energy consumption shown by the PDF in question.

100% reliability will result in a smaller value for E_x . The consequence of reducing E_x is a smaller capacity household battery, and thus a cost-saving for the battery-charging system.

The sensitivity of E_x with respect to the reliability is shown by the shape of the PDF and CDF of daily electricity consumption. For example, if a PDF has a positive skew, this means that there is a long right tail on the distribution. In a CDF, this manifests as a region that becomes increasingly flat on the right. The increase in magnitude of E_x caused by moving from 95% reliability to 99% is relatively high, compared to the increase in E_x moving from 40% reliability to 44%. This means that at low electricity demand values, the sensitivity of E_x (and consequently the required battery capacity) with respect to reliability is high, whereas at high electricity demand values it is comparatively low. Given battery size is correlated with cost, this means that the marginal cost of increasing the reliability rises as the reliability itself increases. (This result was shown to be the case for entire off-grid PV systems through detailed simulation in [39].)

In this way, the shape of the PDF/CDF shows the sensitivity of E_x with respect to the reliability, and so informs the decision of what reliability is logical to set as a target to define E_x . If, as in the example above, a PDF has a positive skew (a long right tail, such that high daily electricity consumption is rare), reducing the reliability (by, for example, 5%) is advantageous for saving cost. A small reduction in reliability leads to a large reduction in E_x and required battery size. However, if a PDF has a negative skew (a long left tail, such that small amounts of electricity are rarely used), a reduction in reliability of 5% results in a much smaller reduction in E_x . Achieving an equivalent reduction in E_x as for the PDF with a positive skew would require a much greater drop in reliability, leading to a higher proportion of unmet demand, and a more significant negative impact on customer experience. Therefore, in the event of a PDF with a positive skew, it is advantageous to reduce the reliability, but if a PDF has a negative skew, the same logic does not apply.

Looking to previous data of electricity consumption, Kwac et al. [116] and Irwin et al. [144] present data for residential Global North contexts, and both data sets show a positive skew. It is expected that similar results will be apparent for rural Global South electricity consumption data. Given the reasoning above, in the event of a positive skew PDF, we set the reliability at less than 100%. The value selected for a particular project or in the design of a particular system to be widely replicated will depend on ensuring the system is affordable for end users and financially viable. In addition, there is no established reliability design point

for this context, though in the concluding remarks of [39] there is a call for greater discussion around setting target reliability standards. For the purpose of enabling comparison between actual numbers in this research, we define E_x as the daily electricity consumption that gives 95% reliability, such that 95% of demand is met. In the event of a negative skew, though this is unexpected, we set E_x as the maximum value of daily electricity demand recorded.

In order to obtain an indicative value for the required electricity demand that must be provided at the hub for a group of households, we define the daily hub electricity demand as the summation of the quantities of electricity used across the households:

$$N_h \sum_{E=0}^{E_x} P(E) \times E \quad (5.4)$$

where N_h is the number of households and $P(E)$ is the discrete PDF of daily electricity consumption. This is essentially the electricity used each day by all the households through the household batteries, which requires replacing at the hub (assuming that it is not possible to use more than E_x energy).

The actual electricity demand at the hub may differ from this daily hub electricity demand because for battery-charging systems, unlike wired systems, the method of electricity distribution is via the portable household batteries. The ways this may occur are discussed in Section 5.4.2.

5.3 Case Study

The case study lighting electricity data set was analyzed to find daily electricity demand distributions for each category of household attribute, and used in the demand estimation tool to calculate electricity demand for a battery-charging hub in a new community. Households with electricity use data for fewer than 8 days were dropped from the data set, because they did not have enough data to be representative of the distribution of daily electricity consumption in the household. This left data for 50 households for analysis.

5.3.1 Demographic Data Acquisition

Surveys were used to gather data about the households using the battery-charging system of the solar nano-grids. Data gathered comprised: the number of occupants in each household; age, gender and occupation of occupants; location and distance of the household to the central

Table 5.1: Availability of household attribute data and correlation coefficient with average daily electricity consumption

Household Attribute	Availability (number of households)	Correlation coefficient with daily electricity consumption
Number of occupants	40	0.36
Average occupant age	39	-0.29
Distance from hub (minutes)	40	-0.39

charging hub; qualitative data about user experience with the system. The qualitative data gathered were not used in this technical investigation but informed evaluation of the project.

For 10 out of 50 households survey data were not acquired: either the households' occupants were unavailable, or surveys were incorrectly labelled so were not attributable to a certain household. Table 5.1 summarizes the household attributes drawn from the survey data, and the number of households for which that attribute is available. As shown in the table, average occupant age was not possible to calculate for one additional household alongside those who did not have survey data. This is because this survey response was incomplete.

With regard to the gender of household occupants, in 17 surveys gender information was not given for 'children' who ranged in age from 0–30 years old. This meant it was not possible to form a household attribute about the percentage of a certain gender in household. Consequently attributes about gender were not investigated in this research.

Table 5.1 also gives the correlation coefficient of a least squares linear regression of average daily electricity consumption on each attribute. All three attributes have a moderate correlation with average electricity use. The strongest correlation is between distance from the hub and daily electricity use, and a similar strength of correlation is apparent between number of occupants and daily electricity use (difference of 0.03 in magnitude of the correlation coefficient). There is a weaker correlation between average occupant age and electricity use. Despite showing a slightly stronger correlation with electricity use, number of occupants was chosen over distance for the analysis because it is easier to measure in a survey. At the beginning of a project, the location of the hub may not have been determined yet, which would make distance to the hub impossible to define. In addition, self-reported time taken to travel to the hub is a more subjective attribute than the number of occupants, as it relies on estimates from end users and may differ depending on the person answering the survey.

Table 5.2: Number of households in each category of Number of Occupants attribute

Category	Number households
1–4 people	17
5–8 people	20
9–12 people	3
Unknown	10

A strong negative correlation was found in a least squares linear regression of average occupant age and number of occupants, with a correlation coefficient of -0.62. The null hypothesis for zero correlation can be rejected to 0.01% significance. This suggests that larger households are also younger households. This negative correlation is not an unexpected result, considering that larger households are likely to contain more children, who bring the average occupant age down. Correlations were also tested between average occupant age and distance (-0.04), and number of occupants and distance (-0.11), and imply weak correlation in both cases.

Three categories were designated for number of occupants that split the attribute into equal segments: 1–4 people, 5–8 people, 9–12 people. There are other ways of forming categories, for example according to the characteristics of the groups: single people living alone, couples living alone, small families, larger families. However, this makes assumptions linking the number of people with a particular living arrangement, the boundaries of the categories are somewhat arbitrary, and similar household set-ups may not be present for other communities. Alternatively, categories could be formed so that there are an equal number of households in each. However, the same distribution of number of occupants may not occur for other communities and so the categories may not be relevant. The chosen way of splitting the attribute into equal segments ensures the same categories across different communities, gives clear boundaries separating each category, and three categories are appropriate to demonstrate this method with results that are easy to interpret. However, it should be noted that this meant that there are just three households in the last 9–12 people category. The number of households in each category is shown in Table 5.2.

5.3.2 Analysis: Demographic Data and Electricity use

Following the methodology in Section 5.2, discrete PDFs and CDFs of daily electricity consumption for the three categories within number of occupants were calculated. In addition,

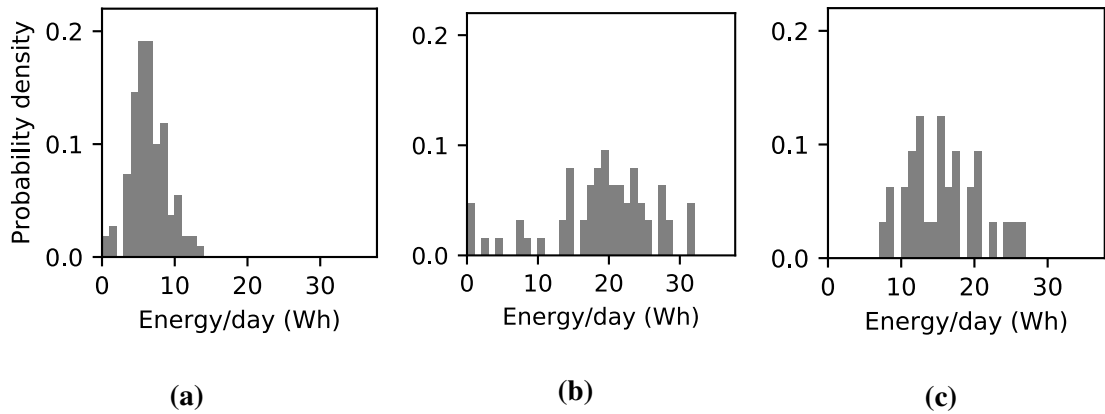


Figure 5.3: Discrete probability density functions of daily electricity consumption for three indicative households: a) low energy household, b) high energy household with negative skew, c) medium energy household.

the nine households that do not have data about number of occupants were put together in a ‘Unknown’ category and the PDF and CDF found.

Figure 5.3 shows the estimated discrete PDFs of daily electricity consumption for three indicative households in the data set. Figure 5.3a shows a household with a low daily electricity use, and b) shows a household with a higher and broader range of electricity use that has a negative skew. Figure 5.3c shows a household with a medium daily electricity use that is approximately normally distributed. The sample of household specific daily electricity consumption PDFs show that there is a range of electricity use behaviours when it comes to the daily electricity requirement across the households.

Figure 5.4 shows the estimated discrete PDFs (left plots) and CDFs (right plots) for each category within the number of occupants attribute. The PDF for a household with 1–4 occupants (Figure 5.4a) is approximately a normal distribution centred at 7 Wh per day. Moving to the higher occupant categories, the shape of the distribution becomes shorter and wider, with a large positive skew for 5–8 people (Figure 5.4b). This implies these households occasionally use higher amounts of energy, but also commonly mimic the smallest households.

For households with 9–12 people (Figure 5.4c), the distribution is less smooth and more erratic, with the same range of values as the previous category, but without the positive skew. The distribution is slightly bimodal, with a collection of daily electricity consumption values at low energy magnitudes at 7 Wh, and another collection around approximately 21 Wh. The bimodality suggests that there is a mode of use that is similar to the low number of

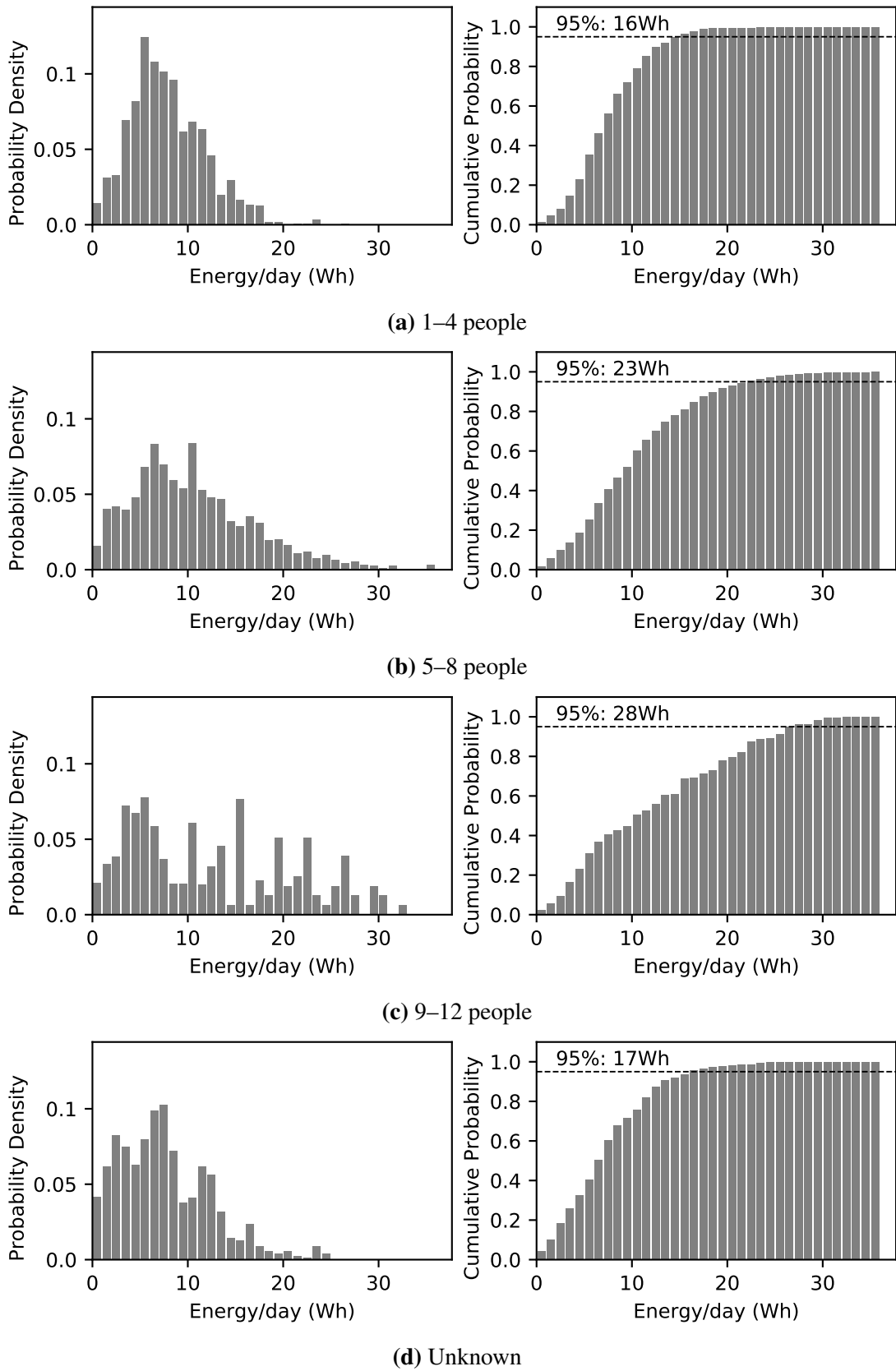


Figure 5.4: Discrete probability density functions (left) and cumulative distribution functions (right) of daily electricity consumption for households with different ranges of number of occupants. On the cumulative distribution functions, the 95th percentile of daily electricity consumption is shown.

Table 5.3: Kolmogorov-Smirnov test results for each pair of household categories. Test statistic ($D_{n,m}$), critical value of the test statistic ($D_{n,m,0.025}$), action taken on the null hypothesis (H_0), and test outcome.

Category Pairs	$D_{n,m}$	$D_{n,m,0.025}$	H_0	Outcome
1–4 and 5–8 people	0.204	0.253	Accepted	Similar
1–4 and 9–12 people	0.341	0.253	Rejected	Different
5–8 and 9–12 people	0.172	0.252	Accepted	Similar

occupants category, but also an additional mode of use associated with larger households where more than double the energy is used.

The PDF for the category of households with no attribute data (Figure 5.4d) is similar in shape to those for households with 1–4 occupants and 5–8 occupants, with a significant positive skew. The Unknown category is most likely made up of households with 1–8 people in.

From visual inspection of the PDFs and CDFs for the different categories in Figure 5.4, it appears that increasing numbers of occupants leads to higher likelihood of higher daily electricity use. The KS test was carried out for each pair of categories, excluding the Unknown category, using equations (5.1) – (5.3)². The test is two-tailed with a significance of 5%. The values for the test statistic, the critical value of the test statistic, the action taken on the null hypothesis, and the test outcome, are summarized in Table 5.3 for each pair.

The results show that only the PDFs for 1–4 people and 9–12 people are statistically different according to the KS test, and it is not possible to reject the null hypothesis for the two other pairs. We expected there to be a difference between all three categories. However, this is not the case for two of the pairs, according to this test.

The CDFs (right hand plots of Figure 5.4) of daily electricity consumption for each category show that E_x for 95% reliability differs across categories. There is a 75% increase in E_x between houses in the smallest and largest categories of number of occupants. These are significant differences in electricity demand that can be used to give an indication of demand for new projects.

5.3.3 Demand Estimation: Community X

The second part of the methodology is a demand estimation tool, which can be used to estimate the electricity consumption for new communities from demographic data gathered

²For each category, the number of observations (n and m in equations (5.1) and (5.2)) was estimated as the average number of observations across the households assigned to that category. In this context, the number of observations is the number of days of electricity consumption contributed by each household.

Table 5.4: Community X demographic data: Number of occupants

Category	Number households
1–4 people	10
5–8 people	15
9–12 people	13
TOTAL (N_h)	38

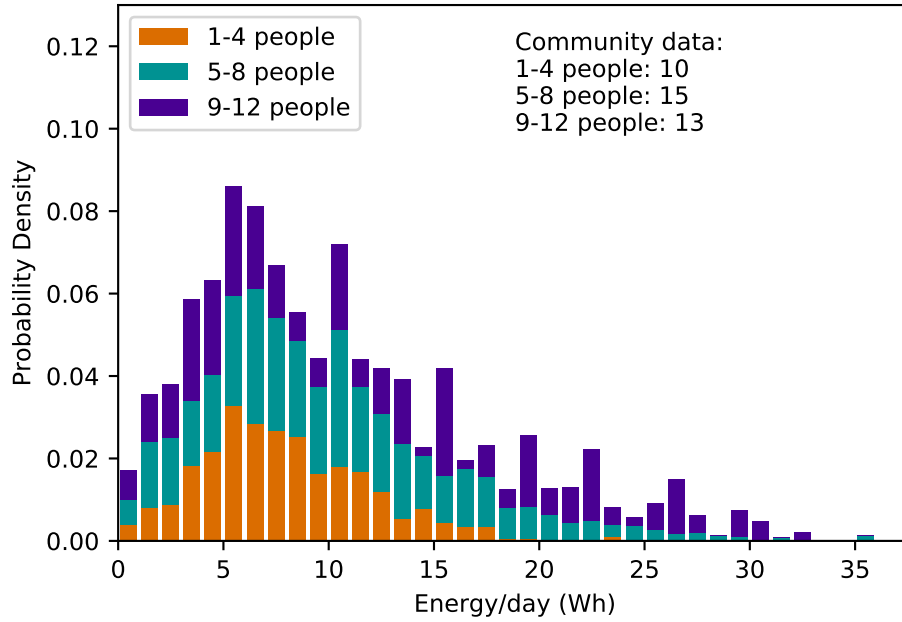
through a survey. It is demonstrated for fictional Community X, using the PDFs for the categories of number of occupants in Section 5.3.2. The data of number of occupants per household is shown in Table 5.4, and is fabricated for the demonstration of the method.

Figure 5.5 shows the discrete PDF and CDF of daily electricity consumption per household for Community X. The shape of the PDF (Figure 5.5a) is a combination of the different categories of number of occupants, indicated by color, in proportion to the percentage of households in the new community in each category. It has a positive skew, contributed by the PDFs for categories of 1–4 and 5–8 people. The highest category of number of occupants contributes a flatter but more erratic shape. The CDF (Figure 5.5b) shows the end-use daily electricity requirement and hub electricity demand for 95% reliability. The hub electricity demand is calculated using equation (5.4), with $N_h = 38$.

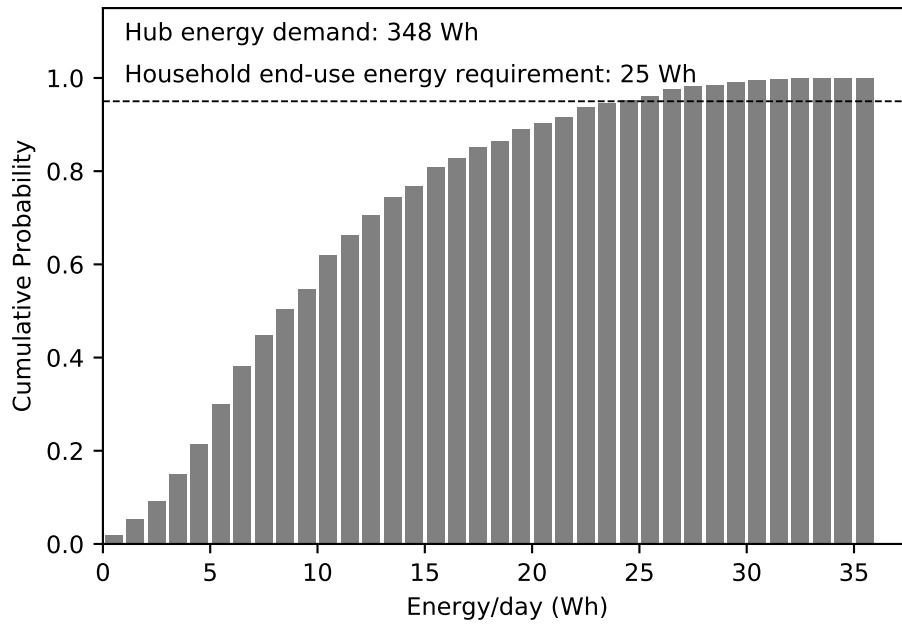
5.4 Discussion

5.4.1 Discussion and Interpretation of Case Study Results

The analysis of the electricity and demographic data from this case study has indicated that there is a correlation between the number of people in the household and the amount of lighting electricity used each day, with larger households more likely to use more electricity. This is interesting because the differences are observed despite the households having the same technology, in terms of the battery and household lighting system. Using the terminology of the Energy Culture's Framework (Section 2.3.2 in Chapter 2), they have the same material culture with respect to the electricity access equipment in their households. In addition, visits to the communities showed that most of the households are similarly sized with two rooms, and surveys showed that most have a similarly low level of existing appliances. The use of the system was also independent from cost in this particular study. Therefore, it is hypothesized that the cause of the different electricity consumption is differences in the norms and practices



(a)



(b)

Figure 5.5: Result of demand estimation tool for fictional Community X: a) discrete probability density function of daily electricity consumption with community characteristics, and b) cumulative distribution function with household end-use electricity requirement, E_x , and hub electricity demand (equation (5.4)).

of households with different numbers of occupants. In the following paragraphs, we draw on observation and personal experience to offer interpretations of these results, suggesting why norms and practices across the households differed.

In larger households, more activities can occur in parallel, with different occupants engaged in more varied activities than are found in smaller households. This may cause use of more than one room at a time or requirements for higher light levels in the same room. In smaller households, fewer activities may be occurring, and essential evening activities such as cooking and cleaning may occur sequentially rather than in parallel. In addition, activities that scale with the number of household occupants will take longer in larger households, which may also contribute to higher electricity use. For example, cooking and ironing will take longer in large households because there are more people to feed and more clothes to iron.

An additional explanation that is supported by the results of the case study is that the different electricity use is influenced by more children being present in larger households. The evidence for this is the strong negative correlation between average occupant age and the number of occupants observed for this data set. These children may be attending school and want to study in the evenings, which requires task lighting to be able to see to read and write. Occupants in smaller households containing only adults may not have wants or expectations around studying, and so have fewer practices that require task lighting compared to background lighting. There are likely many other ways that the presence of children lead household occupants to have different norms and practices - for example, changing expectations around entertainment and care-giving duties.

The definition of E_x was set to achieve 95% reliability in terms of the ability of the household battery capacity to deliver electricity demand each day because a positive skew in electricity use was observed. This was most evident in the PDF shape for households with 5–8 people and the households with Unknown demographic data, though a less strong positive skew is also observable in the other household category PDFs. This is therefore a good design point to set, because for a small reduction in reliability, a large reduction in the size of equipment is achieved. For community X, by taking E_x for 95% reliability rather than 99.9% the energy capacity of the household battery required is 31% lower, which would translate into a large cost-saving. In electricity access projects, where the end users have very limited money to spend on electricity, this cost-reduction could be passed on to the end users, increasing affordability of the project. However, the results from this case

study show that larger households are the demographic who more frequently use more than 25 Wh a day. This approach would disproportionately affect higher occupancy households over low occupancy households, who would not be able to meet their electricity needs on higher demand days. These households are also the youngest households, with the most children. The issue of cost reduction through reducing the percentage of household demand met, compared to the problem that this adversely affects larger households with children, would be a key trade-off in design decision making.

Use of the household attribute category PDFs derived for the case study to predict electricity consumption in other villages should occur with an understanding of the assumptions involved. The issue of data transferability and associated assumptions are discussed in Chapter 4, Section 4.5. However, given the lack of publicly available demand data for tier 1 users, they can serve as a first demand estimation in project development, or for further research in electricity access systems.

5.4.2 Sizing Key Parameters for Battery-Charging Systems

This section discusses the factors involved in moving from the household end-use electricity requirement, E_x , and hub electricity demand calculated in Section 5.3, to sizing the household battery and charging hub generation for a battery-charging system.

5.4.2.1 Household Battery

Determining the appropriate size (energy capacity) of the household battery depends on how much electricity is required in the household, how long the battery is required to last between charges, and the depth of discharge factor of the battery. In this analysis, E_x is set at the 95th percentile of the CDF of household demand, and this gives a basis from which to size a battery. The depth of discharge is simple to factor in to battery sizing. It depends on the battery chemistry, and the extent one wants to oversize the system to prolong battery life. Recommended values can be found on manufacturers' documents. The depth of discharge increases the required battery capacity by a factor that is the inverse of the depth of discharge in decimal form.

Having determined daily electricity use and depth of discharge factor, the total battery energy capacity depends on how long batteries are to last between charging. If batteries were to be charged every day, E_x modified to include the depth of discharge factor would be the

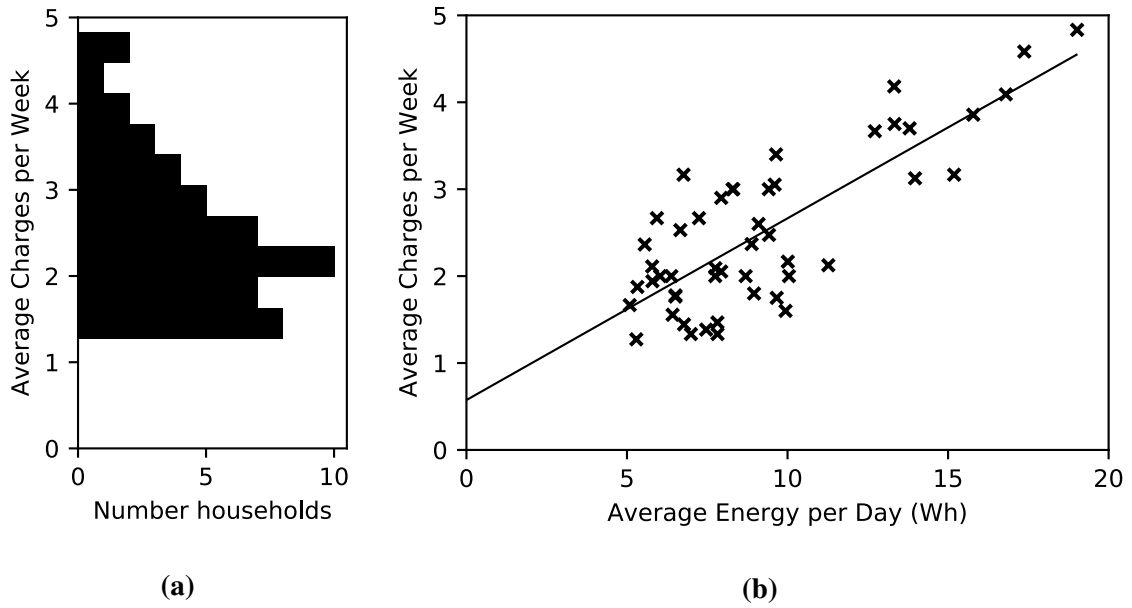


Figure 5.6: Average number of times households charge per week, a) histogram and b) plot against average daily electricity consumption. Line of best fit on b) calculated by ordinary least squares.

battery energy capacity. However, the feedback from the case study strongly suggested that all users thought larger batteries were desirable: for the obvious reason that less frequent charging would reduce the number of inconvenient visits to the charging hub. The actual data show a range of charging behaviour in the case study: see Figure 5.6a, which is a histogram showing the average number of charges per week for the households. (The size of the household batteries in the case study was 62 Wh.) No households charged an average of every day, and the range of charging behaviour is from once to five times a week, skewed towards charging less frequently.

As one might expect, charging frequency is correlated with electricity use – such that the more electricity is used, the more often the battery is charged. This intuitive result is shown by Figure 5.6b, where average charges per week is plotted against average daily electricity consumption for each household. Ordinary least squares is used to calculate a best-fit line. The p-value for the null hypothesis for no correlation is zero to 10 decimal places, so the null hypothesis is clearly rejected. The correlation coefficient is 0.8, indicating a strong positive correlation between the average number of charges per week and average electricity used per day.

The feedback suggests that users want to charge as infrequently as possible, and so the bigger the battery, the better for user satisfaction. Therefore, one approach to sizing the

battery is to consider the financial costs, as well as any other operating limits that might come into play as battery size is increased. Larger batteries are more expensive, and the increased cost would typically be passed on to consumers. There is likely to be a point at which the size of the battery is such that the system becomes too expensive for users (how the sensitivity of usage to price can be determined is discussed below). There may also be a portability limit; as batteries increase in energy and power capacity, they become physically larger and heavier (for the same battery chemistry), and depending on the chemistry, there will be a point at which they become difficult to transport to the charging hub. The point at which both of these limits are reached will vary across and within communities due to a range of socio-economic circumstances, geographical factors and physical ability.

Keeping options within the portability limit, the optimal battery size from a cost perspective can be established from a ‘willingness to pay’ survey, by presenting people with various options. For example, various costs of different battery sizes, along with associated battery time between charges given the estimated electricity use. The results would be expected to depend on a number of factors, including the financial circumstances of the community.

However, an important caveat to the logic of the approach in this chapter is the assumption that electricity consumption will not change if the technology changes, for example, if the battery size changes. It is assumed that the output of the analysis presented is the electricity requirement of the users – hence the terminology, ‘household end-use electricity requirement’ for the electricity demand selected for design. There is a weakness in this assumption, because in fact, the recorded levels of electricity use in the case study will have been determined in conjunction with the charging frequency by each household, in response to a trade-off. The trade-off is between the importance or benefits of using electricity, and the inconvenience of charging, which is the time and effort cost of using electricity for the battery-charging system. (In the thesis case study, the financial cost was independent of the electricity used, but in practice this might not be the case and cost will be an important consideration).

The electricity data analyzed here are the electricity use and charging frequency operating points that result from this trade-off, between the benefits and costs of electricity use. Although the feedback from the case study suggests that larger batteries are important to end users, to reduce the number of times they need to charge, there is no reason to believe (on the basis of the above arguments), that a household’s electricity consumption would remain the same in response to a change in battery size, with only the charging frequency affected. Instead, it

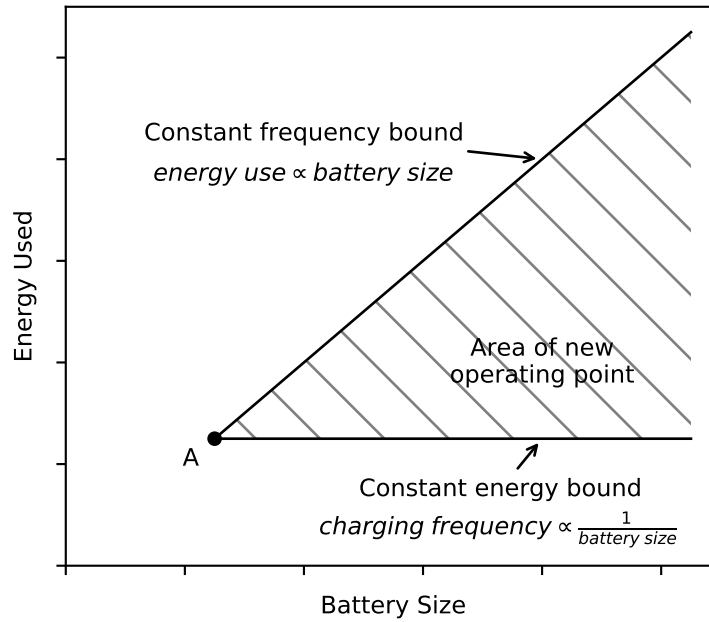


Figure 5.7: Schematic showing how electricity consumption is expected to change with battery size. Starting from operating point A, two bounds of constant energy consumption and constant charging frequency are marked. Actual energy consumption with respect to battery size is expected to be in hatched area.

is more likely that the relationship between electricity consumption and charging frequency would change for a household. The expected situation is shown in Figure 5.7, which shows the electrical energy used against battery size. Point A is the current operating point in terms of energy used and charging frequency for the given battery size. Two bounds are marked that show the extremes of how the energy used can be expected to change with increasing battery size. The constant energy bound (horizontal from A) represents the assumption that changing the battery size results only in a change in charging frequency. As battery size increases, energy use stays the same and users charge less frequently. The constant frequency bound represents the other extreme, where an increase in battery size results in an equivalent increase in daily energy use, and charging frequency stays the same. We argue that in practice, neither of these bounds would describe the relationship between battery size and energy use, and that both charging frequency and energy use would change, with a new operating point within the hatched area. This is because an increase in battery energy and power capacity would likely be taken advantage of by the end users to run more lights, for longer, and (unless constrained by the system) more appliances. It is argued that a larger battery would reduce charging frequency, but also increase electricity use, such that the relationship between electricity use and charging

frequency is itself a function of battery size. The end-use electricity requirements in this analysis should therefore be used with awareness that battery size may affect the magnitude of electricity used in the home. However, used in the sizing strategy described above they give an indication of how long a given battery size will last between charges.

5.4.2.2 Hub Generation Capacity

The daily generation capacity at the hub depends on the electricity demand, set by the number and size of batteries charging each day, and system losses. System losses increase the daily electricity demand at the hub. One source of losses is charging losses associated with the household batteries, caused by internal resistance of the batteries. This means a portion of energy entering the batteries is lost as heat during the charging process. In addition, there are losses in the battery charger, in the distribution wires, and any control equipment present such as a voltage regulator. The size of expected losses can be obtained from manufacturers specifications of equipment.

In this analysis, the hub electricity demand (equation (5.4)) is an indicative value for the electricity required at the hub (ignoring losses), which is the summation of the energies consumed in each household in a day according to the PDF. For a wired electricity access system, this is an accurate indication of the average electricity required by the households for a day. However, the way a battery-charging hub works means that the actual electricity demand at the hub will fluctuate, depending on the number of batteries present for charging. It is not necessarily the case that there will be an equal number of batteries present at the hub each day. A detailed investigation and simulation of how the number of batteries present at a battery-charging hub varies is not within the scope of this thesis, but a simple analysis of charging frequencies is possible that gives some indication of what would occur in reality.

Figure 5.8 shows the number of batteries that would be present at a charging hub throughout a year, using data extracted from the case study. The simulation uses the average charging frequencies for users in the case study, and these are assumed constant throughout a year for 50 users. (This approach, rather than using the raw time series data for the case study, simulates a situation where the number of batteries in the community is not fluctuating and decreasing due to technical issues). The graph shows that after the first day when all 50 batteries are charged, the number of batteries present fluctuates around a mean value. Due

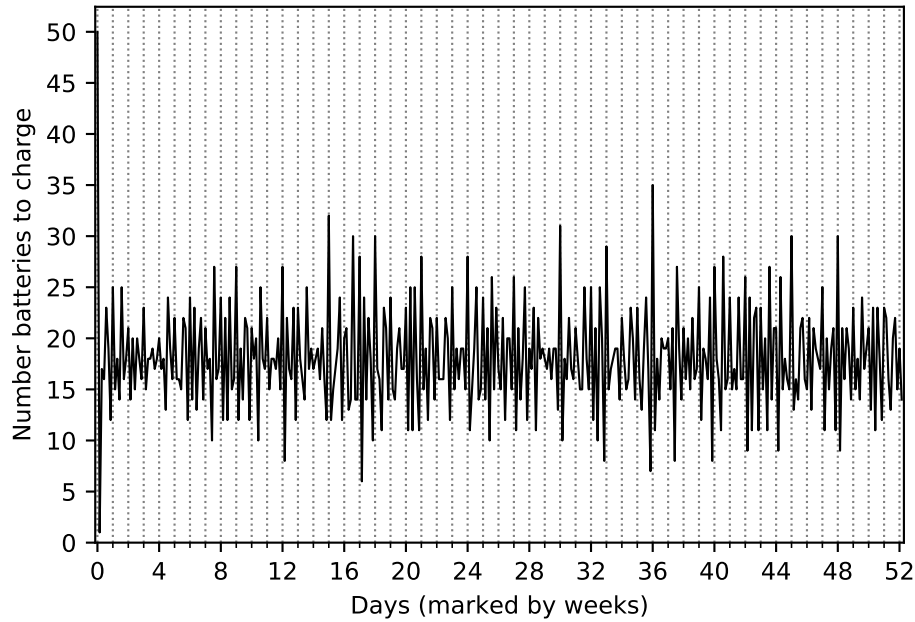


Figure 5.8: Number of batteries present at a charging hub, for 50 households with constant charging frequencies, based on case study data. Weeks marked with dotted lines.

to different charging frequencies across users, the number of batteries to be charged each day varies, which means the electricity demand at the hub also varies.

The situation presented in Figure 5.8 is a simplified case, because it assumes each user has a constant charging frequency throughout the year. In fact, charging frequency is more likely to vary for each user, and could be modelled using a gaussian distribution centered on the mean charging frequency. In addition, norms and social factors may affect users' practices about charging. For example, different routines may occur on particular days of the week, or throughout the year, in response to social holidays, the rhythm of the working week, or less predictable events like the weather. How both of these aspects affects the number of batteries present at the hub is out of scope of this research. However, we estimate that the factors would work in opposite directions: the variance in charging frequencies will give rise to diversity in the situation shown in Figure 5.8, smoothing extremes out; the social factors will encourage similar practices across a community, leading to more batteries present on certain days.

This short analysis has shown that the number of batteries at a charging hub would vary daily, unless manual intervention occurred to schedule batteries evenly across days, or there was a physical limit due to the number of charging ports (either of which may cause customer dissatisfaction). Therefore, the hub electricity demand (equation (5.4)) is an indication of electricity required at the charging hub, but we are aware that it is dependent on the variation

in the number of batteries present each day. Despite this, it is a more pragmatic value of hub electricity demand than that derived by assuming that provision is required for all batteries to be charged during one day. Designing to this criterion would ensure that the charging hub was always able to meet the electricity demand from the batteries. However, for the majority of the time, the system would be operating well below the loading specification for which it had been designed, and the extra cost of the generation and charging port components would impact the financial viability of the system. This most extreme loading scenario is unlikely to occur frequently in reality, and when it does arise, it could be managed to reduce energy and power requirements. Ultimately, the extent to which the generation and charging infrastructure at the hub could be over-specified to cope with more batteries above an average baseline would depend on system costs, and the effects on affordability for end users and profitability for service providers.

Further detailed modelling of the number of batteries expected to be charged each day, variation in this, and the impact of variance in solar irradiation, is out of the scope of this research, but is possible through the electricity data provided in Chapter 4.

5.5 Summary

This chapter addressed the question: How can electricity demand and demographic data be used to estimate daily electricity consumption?

A method to use measured electricity and demographic data was outlined, to first link daily electricity consumption to demographic data, and then estimate electricity demand for a new community from demographic data. The approach was taken to use discrete probability density functions (PDFs) of daily electricity consumption, to retain the variation in electricity use within and across households. The method shows how PDFs for different types of households (according to a single demographic variable) are found, which can then be used to estimate the PDF of electricity use for a group of households from demographic data. This shows the range of electricity consumption estimated for the new community, from which design points of daily electricity use, required to size the household battery, can be inferred. From the PDF, the electricity demand at the hub can also be estimated. In the discussion, we also explore the factors that contribute to the translation of the demand calculations into sizing of battery-charging hub components.

What are the typical patterns of electricity use within and across rural, low-income users?

6

Typical Electricity Use Patterns of Tier 1 Users of Rural Nano-grid

Contents

6.1	Introduction	139
6.2	Methodology	141
6.2.1	Clustering Analysis Stages and Evaluation Criteria	141
6.2.2	Application of Clustering Analysis to Time Series Electricity Data For Each Clustering Stage	144
6.3	Case Study Results	147
6.3.1	Representative Load Profiles for Households	148
6.3.2	Communal Energy Modes Across Households	155
6.4	Discussion	161
6.5	Summary	165

6.1 Introduction

The previous chapter focused on electricity demand in the form of total daily electricity consumption, which is an important metric for sizing components in battery-charging systems. The two steps in the approach to demand estimation for new communities, of 1) analysis of existing electricity data to understand electricity use, and 2) linking electricity use to demographic data, were presented. In this chapter and the next the focus is electricity demand in the form of load profiles that show the magnitude and time of use throughout the day. Load profiles are a richer form of electricity data, and require a more complex process of analysis

to understand how electricity is used for a particular context. This analysis step is presented in this chapter, and in Chapter 7 the link to demographic data, and subsequent load profile estimation for new communities, is outlined. The purpose of the analysis in this chapter is to extract the typical patterns of electricity use within and across rural, low-income users.

Clustering analysis, reviewed in Section 2.3.1 of Chapter 2, is used on time series electricity data of rural, low-income users to extract a concise set of typical daily use patterns. As identified in the gap analysis in Chapter 2, Section 2.2, existing work in load profile classification models assigns one representative load profile (RLP) to each customer. This does not take account of each customer's variation in electricity use between different days. Therefore, the approach taken in this analysis retains the variation of use for each customer, using a two-stage clustering approach as in [102], which has not previously been applied to Global South electricity data. The clustering process is performed twice. After the first stage, each household is represented by multiple 'household RLPs', which show the typical daily patterns of electricity use for a particular household, and how common they are. In the second clustering stage, the household RLPs for all households are compared to each other and grouped according to similarity. The clusters formed from this stage describe how the household RLPs vary across the data set, and are called 'communal energy modes'. This approach reveals more detail about the use patterns within a community, and does not reduce each household's use to an average pattern, which is not necessarily representative of the actual modes of use that are occurring. This detail in particular is useful to show the higher energy and power modes of use, which may otherwise not be visible if average use patterns were used.

The process to apply the clustering analysis to Global South electricity data is described in the Methodology section (6.2). At each stage of the clustering, the steps to carry out the analysis are the same. Just the input data (days or household RLPs), and the purpose of the clustering, differs between the two stages. Therefore, the clustering steps are described so as to be applicable for each clustering stage. In section 6.3, the results for the case study are presented. This is split into two sections to show each stage of the clustering analysis, showing first the household RLPs, and then the derivation of communal energy modes. For each of these sections, we first outline how the particular clustering solution was chosen, and then present the accompanying results. Finally, in Section 6.4, the results are discussed in terms of the insight they give of electricity use in tier 1 households using

a battery-charging system, the diversity of use patterns within and across households, and implications for electricity access system design.

6.2 Methodology

The range of clustering algorithms, and methods to choose a ‘best’ clustering solution, were reviewed in Chapter 2, Section 2.3.1. Given the choice of clustering algorithm influences the result of the analysis, the approach is taken to use two clustering algorithms that have different behaviour, and set out a process to choose the best result from them. K-means and Hierarchical agglomerative clustering (HAC) are selected, as k-means works best with convex, isotropic clusters, while HAC can work with non-convex, anisotropic clusters. HAC has four options for linkage criteria, and these provide an additional range of solutions. (See Tables 2.2 and 2.3 in Chapter 2 for a review of the clustering methods and HAC linkage criteria).

To choose a clustering solution, we primarily follow the ethos of von Luxburg et al. [126]: interpreting the results in light of the context of the analysis and what the results will be used for. This requires setting out the criteria with which to judge a good clustering result, which are different for each stage of the two-stage clustering analysis. Therefore, these are described alongside the method of the two-stage analysis in the following section. As a secondary input to inform choice of a clustering solution within a particular clustering algorithm (rather than to choose between algorithms), an appropriate internal quality score is used. This is described alongside the other steps to carry out the clustering analysis for each stage in Section 6.2.2.

6.2.1 Clustering Analysis Stages and Evaluation Criteria

Figure 6.1 summarizes the steps of the two-stage analysis used to understand the typical patterns of electricity use within and among the households. In stage 1, for each household individually, household RLPs are found. For each household, the days of electricity use are compared to each other and grouped according to similarity. The centroids of the clusters formed are the RLPs for that household, and each one represents the days of electricity use in a specific cluster. If present, the most abnormal days (outliers) are identified and analyzed (see Section 6.2.1.1).

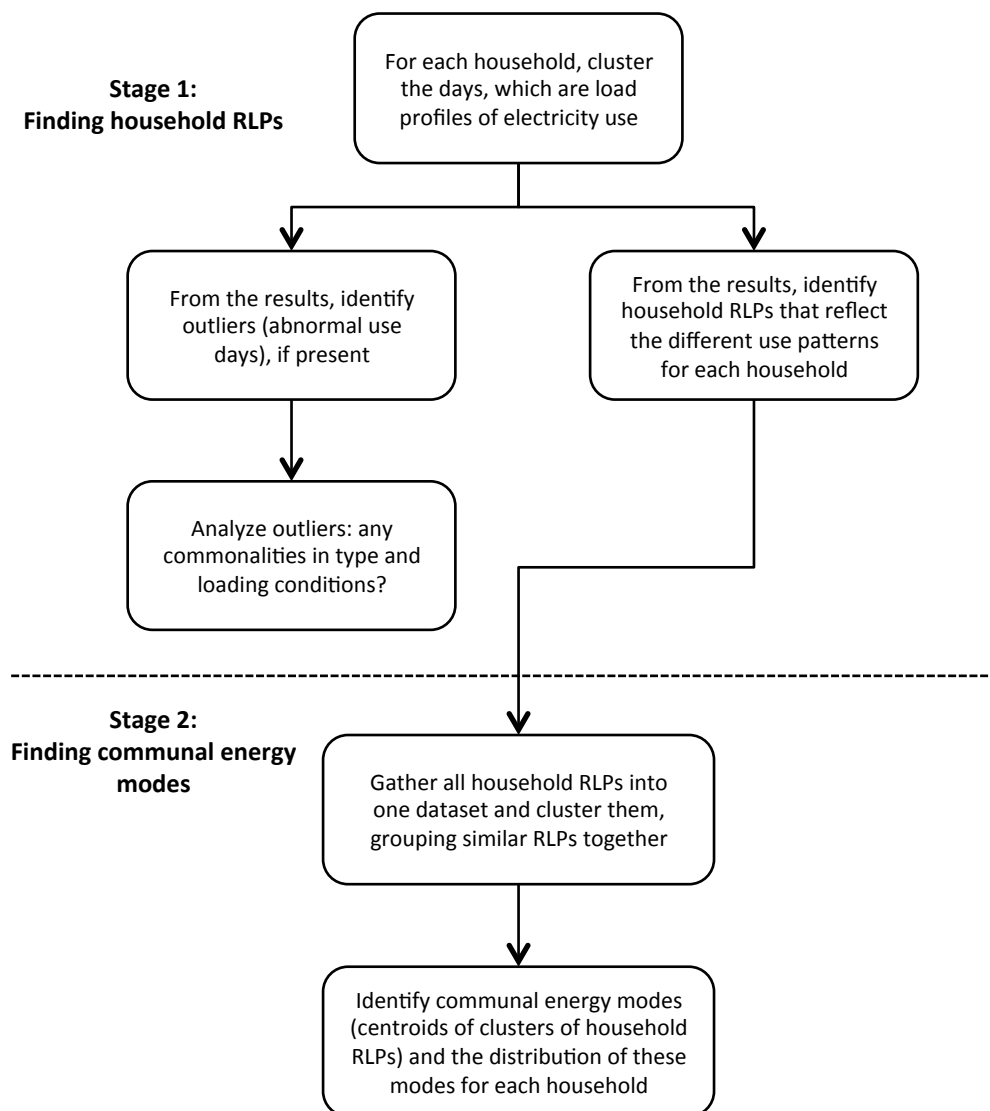


Figure 6.1: The two-stage clustering analysis for time series electricity data, to first find household RLPs (representative load profiles), and then find communal energy modes.

In stage 2, the household RLPs for all households are clustered together and grouped according to similarity. The centroids of the clusters formed are communal energy modes, and each represents the group of household RLPs that make up a particular cluster.

In both cases, household RLPs and communal energy modes are the centroids of the clusters formed as a result of the clustering analysis. An alternative approach is to take the medoid to represent the cluster, which is the closest actual data point to the cluster centroid. However, for small clusters (for example, of less than 10 data points), the medoid is not necessarily representative of all the data points in the cluster, as all the points may be on the ‘edge’ of the cluster. This is relevant for some households in the case study because of the low number of days of electricity use they have for clustering.

6.2.1.1 Identification and Definition of Outliers

Some clustering algorithms and solutions might identify single or small groups of data points as clusters. If these are apparent in the clustering of the raw days of electricity use (stage 1), these are interpreted as use days that are abnormal in terms of time of use and/or magnitude of use. They represent extremes of use, and therefore may not be appropriate to be included in analysis of typical use patterns. Given this, we term them ‘outliers’. However, outliers may be interesting to explore separately, for example, by examining the type of outliers that occur and any links with specific loading conditions. They may still contain useful information about the electricity use for the design of the electricity system, or for understanding electricity behaviour.

If a clustering result does give rise to small groups or individual days, and their definition as outliers is meaningful in the problem context, a definition must be set that determines whether a cluster is an outlier, or big enough to be representative of a typical use pattern. The approach taken in this research is to set a minimum ‘valid’ cluster size equal to a percentage of the total days in the data set. This percentage is defined by examining the clustering results, and in conjunction with choosing the clustering solution (number of clusters), as the number of small clusters labelled and eliminated as outliers will depend on the cut-off percentage. Where outliers groups exist, we define N as the number of actual clusters found for a solution with a total of k clusters (giving $k - N$ outliers).

As discussed in Section 2.3.1 of Chapter 2, different algorithms will have different properties in terms of the identification of individual data points that are outliers in this context. For example, of the hierarchical linkage criteria, single linkage is likely to find the most outlier days given its tendency to chain samples onto existing clusters. Ward’s linkage criteria is less likely to identify single member or very small clusters, because it creates more equal sized clusters by minimizing the total inertia at each step.

6.2.1.2 Evaluation Criteria

The criteria for a ‘good’ clustering solution are defined for each clustering stage, so that clustering results can be judged against them and the best solution picked.

For stage 1 (extracting household RLPs), the aim is to uncover the typical use patterns for each household while accurately capturing the variation in the data. Therefore, if there

are abnormal use days (outliers) that do not represent typical use, these should be identified. Ideally, a concise set of household RLPs should be identified alongside potential outlier days. This means the solution sought is one that has a range of cluster sizes, picking out the most different electricity use days (outliers) as single-member or very small clusters, and providing a grouping of the remaining days in larger clusters so that they can be identified as representative load profiles. That the clusters derived are all equally sized is not important, because if one pattern is more common than another, this is useful information that the second stage aims to draw on.

For stage 2, extracting communal energy modes, the aim is to derive a concise set of typical patterns that describe electricity use across the data set. In this case, identifying outlier household RLPs is not interesting, because this stage seeks to understand the trends in the data rather than examine specific households. Therefore, a clustering solution that gives a small number of relatively equally sized clusters is preferable.

6.2.2 Application of Clustering Analysis to Time Series Electricity Data For Each Clustering Stage

For each clustering stage, both for stage 1, where the daily electricity load profiles are clustered (in a separate exercise for each household), and stage 2, where household RLPs are gathered together from all households and clustered in one exercise, the following steps apply.

Let a set of load profiles be arranged as a matrix, $\mathbf{L} \in \mathbb{R}^{n \times p}$, consisting of n load profiles, each of which is a vector of p average power values. Pre-clustering steps are applied to $\mathbf{L}$ as described in the following sub-section, to produce a transformed matrix of load profiles $\tilde{\mathbf{L}}$. The clustering algorithms are applied to $\tilde{\mathbf{L}}$, and results viewed in relation to the original data matrix, $\mathbf{L}$.

The objective of the clustering is to find k groups/clusters of load profiles, $S = S_1, S_2 \dots S_k$. The centroid of each cluster is a vector of length p , formed by taking the average value of each feature over the members of the cluster in question. For a particular clustering solution, the centroids form a set, $c = c_1, c_2 \dots c_k$.

For the first stage of clustering, the above is repeated for each household as a separate exercise, and the input load profiles $\mathbf{L}$ are the days of electricity use associated with the particular household. For the second clustering stage, $\mathbf{L}$ are the household RLPs, which are the centroids of the clusters formed for each household, carried forward from the first analysis.

In order to achieve this clustering analysis, the same three steps are applied for each stage, described here. These steps are described so as to be relevant for the clustering of time series electricity data, using k-means and HAC, in general.

6.2.2.1 Step 1: Pre-clustering Steps

If comparison of the shape of electricity use is important, and comparison of the magnitude of electricity use is of less interest, the data should be normalized. If magnitude is important in the analysis, normalization should not be used.

Before clustering, principal component analysis is carried out on the set of load profiles $\mathbf{L}$ to transform it to $\tilde{\mathbf{L}} \in \mathbb{R}^{n \times q}$ with $q < p$ (see Section 2.3.1.2 in Chapter 2). The number of components q is selected to contain at least 95% of the variance. However, this depends on the amount of variation in the data and the number of components required to represent it. If there is a great deal of variation in the data, it may be necessary select fewer components at the cost of less of the variation captured.

6.2.2.2 Step 2: Application of K-means and HAC

K-means is initialized using the k-means++ variation, and multiple values of k are tested and the results recorded. For HAC, the clustering is run for each linkage criterion. A range of k values are then investigated, with a sensible range to give a concise number of clusters being $k = 2 - 7$. Higher k can be investigated if there is lots of variation in the data, and visualization of the results still shows distinct centroids are being generated. In addition, if there are outliers, higher k values may be necessary to examine to extract a set of N typical use patterns, if intermediate values are merely identifying outliers.

6.2.2.3 Step 3: Evaluation of Clustering Results to Choose a Solution

To choose between the different clustering algorithms and variations, and to choose the number of clusters within that, emphasis is placed on visualization of the results and evaluation against the criteria in Section 6.2.1.2. To do this, the centroids of the resulting clusters for each algorithm and k value, and the associated heat maps are plotted. The heat maps should be presented so that the days, or household RLPs, in each cluster are grouped together. Inspection of these can show the k value where the additional cluster formed does not seem significantly different from other clusters/centroids. Depending on the trade-off between obtaining a simple

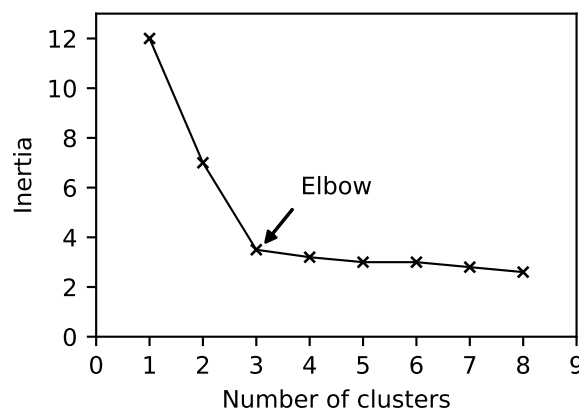


Figure 6.2: The elbow method to determine optimum number of clusters k for k-means clustering. The graph shows inertia of each clustering solution plotted against the number of clusters, showing a clear elbow at $k = 3$ (mock data).

clustering solution with an easily interpretable number of clusters, and capturing adequate diversity in the data through a high enough number of clusters, a solution that makes sense for the context of the analysis can be chosen.

For k-means, an internal clustering quality score called the inertia is also examined for each k , because it reflects the objective function the algorithm seeks to minimize. The inertia of a clustering solution is the total within cluster sum of squares, as defined by equation (2.4) in Chapter 2. Looking for where the inertia drops rapidly for additional numbers of clusters, or conversely, where little difference in inertia occurs, is called the ‘elbow method’. The inertia is plotted for each k that has been tested. The plot is inspected for a sharp bend, which represents the k at which adding another clustering does not correspond to a significant decrease in the inertia of the solution. An example plot with mock data is shown in Figure 6.2, which shows inertia decreasing with increasing number of clusters, and a clear elbow at $k = 3$. It should be noted that the minimum inertia is found when k is equal to the number of data points, so taking the solution with minimum inertia is not a useful clustering outcome.

For comparing different values across k for the same data set, the elbow method is a useful graphical heuristic that tracks the change in the objective function of the algorithm. However, unless there is a clear underlying classification in the data which k-means can find, real data are unlikely to have such a clear elbow as demonstrated in Figure 6.2. It is possible there might be one or more vague elbows or none at all, particularly in this case of exploratory data analysis where a ground truth of classification is not necessarily expected and may not exist. The elbow method should be used in conjunction with visualization of the results, as described above.

6.3 Case Study Results

The heat maps of household lighting electricity use for the case study in Appendix C show that the majority of electricity use occurs in the evening. In addition, they show that electricity use in the morning varies between households and is sporadic and of low magnitude when it occurs. Therefore, the analysis focused on evening electricity use patterns, defining a ‘day’ as between 15:00 and 03:00 for each daily profile. The 15:00–03:00 period was defined because although most households stopped using electricity by midnight, the heat maps show that eight households had four or more days, and 4 households had 10 or more days, where their electricity use extended into the early hours of the next day, and so extending the time span captured the vast majority of regular evening electricity use. The resulting daily load profiles consisted of 24 half-hour periods such that $p = 24$.

Households with fewer than 8 electricity use days, defined as a 15:00–03:00 period where electricity was used in the household (as opposed to when the battery was being charged, or was not in use), were discarded. This is because a week or less worth of data is not sufficient to indicate different patterns of use between the days. This left 49 households in the data set. For all clustering operations, given the low magnitude of average power values in the data, and the fact that comparison of the magnitudes was of interest, the input data set was not normalized.

When PCA was used to reduce the dimensions of the data for each clustering analysis, a judicious choice was made of $q = 10$ to transform each set of load profiles such that $\tilde{\mathbf{L}} \in \mathbb{R}^{n \times 10}$. For the 49 households in the first clustering stage, and for the analysis in the second clustering stage, this contained a minimum of 97% of the variance of the data for each data set. Figure 6.3 shows how the percentage of the total variance of the data captured increases with increasing number of components for an indicative household, and that 10 components occurs after the variance has levelled off.

The results are presented for each stage of the clustering analysis separately. The centroid plots are presented as line graphs, though the value for each half-hour segment applies over the whole half-hour segment. The line graph format has been chosen for ease of interpretation of the use pattern represented.

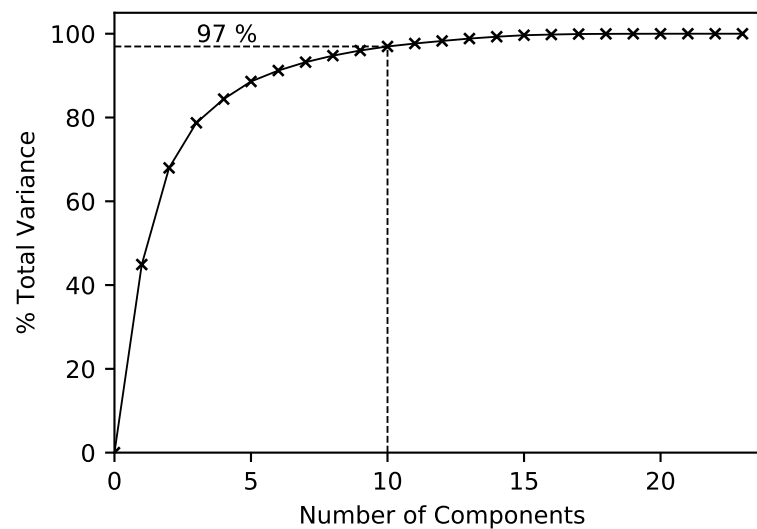


Figure 6.3: The percentage of the variance in the data captured as the number of PCA components selected increases, for an indicative household (household 41). 10 components contains 97% of the variance.

6.3.1 Representative Load Profiles for Households

For each of the 49 households, household RLPs were found in the first clustering stage. The approach was taken to find one clustering solution, comprising clustering method and number of clusters, that was judicious for all household profiles rather than attempt individual optimization of each household profile.

6.3.1.1 Clustering Method Selection

The methodology in Section 6.2 was used to investigate k-means and HAC with four linkage criteria. Due to space constraints, a summary of the results of the algorithms and the reason for selection of the final clustering solution are given here, rather than the result of each stage of the method in Section 6.2.

Figure 6.4 shows the outcomes of the five different clustering algorithms as heat maps with days grouped together according to how the algorithm groups them for an example household. The first four heat maps are for HAC with the four different linkage criterion for $k = 7$, and the fifth plot shows the result for k-means $k = 3$. The different linkage criteria for HAC show a range of clustering approaches on a spectrum between isolation of the most different days and small groups of days only (single linkage), to grouping days in comparably sized groups like k-means (complete and Ward's linkage). An intermediate approach is seen

with average linkage, where some individual days and very small groups of days are isolated, and the remainder of days are split into larger groups of varying sizes. The k-means algorithm splits the data into relatively equally sized groups.

The result provided by HAC with average linkage is meaningful because it is able to distinguish the most different use days and reasonably group the remainder. This is important for finding household RLPs, because it removes outliers that are not representative of typical use, but still finds distinct groups from the remaining days. The outliers were then analyzed further to see if they were important to consider in system design, with results presented in the next section.

Clusters were defined as outliers if they were smaller than 5% of the total number of days in the data set for a particular household. The k value at which a certain number, N , of non-outlier clusters occurred was different for each household according to how varied use was in that household. Inspection of the centroids and heat maps of various N values showed that $N = 2$ and $N = 3$ captured the diversity in each households' daily electricity patterns well. Figure 6.5 shows the household RLPs for an indicative household for both N values (outlier days not shown). The household RLPs are differentiated by both time of use and magnitude of use. For the $N = 2$ solution (Figure 6.5a), RLP A represents days where electricity use occurs earlier in the evening, whereas RLP B represents higher magnitude electricity use later in the evening. For the $N = 3$ solution (Figure 6.5b), an additional low use cluster has appeared. The difference between the two N solutions is that two of the RLPs for $N = 3$ (A and C) are agglomerated to create RLP A for the $N = 2$ solution. RLP B stays the same. Both $N = 2$ and $N = 3$ solutions for all households were carried through the two-stage clustering process to the end. However, the additional information gained from the $N = 3$ case is minimal and so the results are not presented and discussed in this thesis. From this point we show only the results for the $N = 2$ case.

6.3.1.2 Clustering Results

Figure 6.6 shows the household RLPs and heat maps for $N = 2$ for an additional two households. The plots show examples of where either difference in time or difference in magnitude is the dominant factor in the differentiation of clusters. For household 48 (Figure 6.6a), the largest cluster represents earlier evening electricity use, whereas the second cluster represents use that stretches later in the evening and peaks later. For household

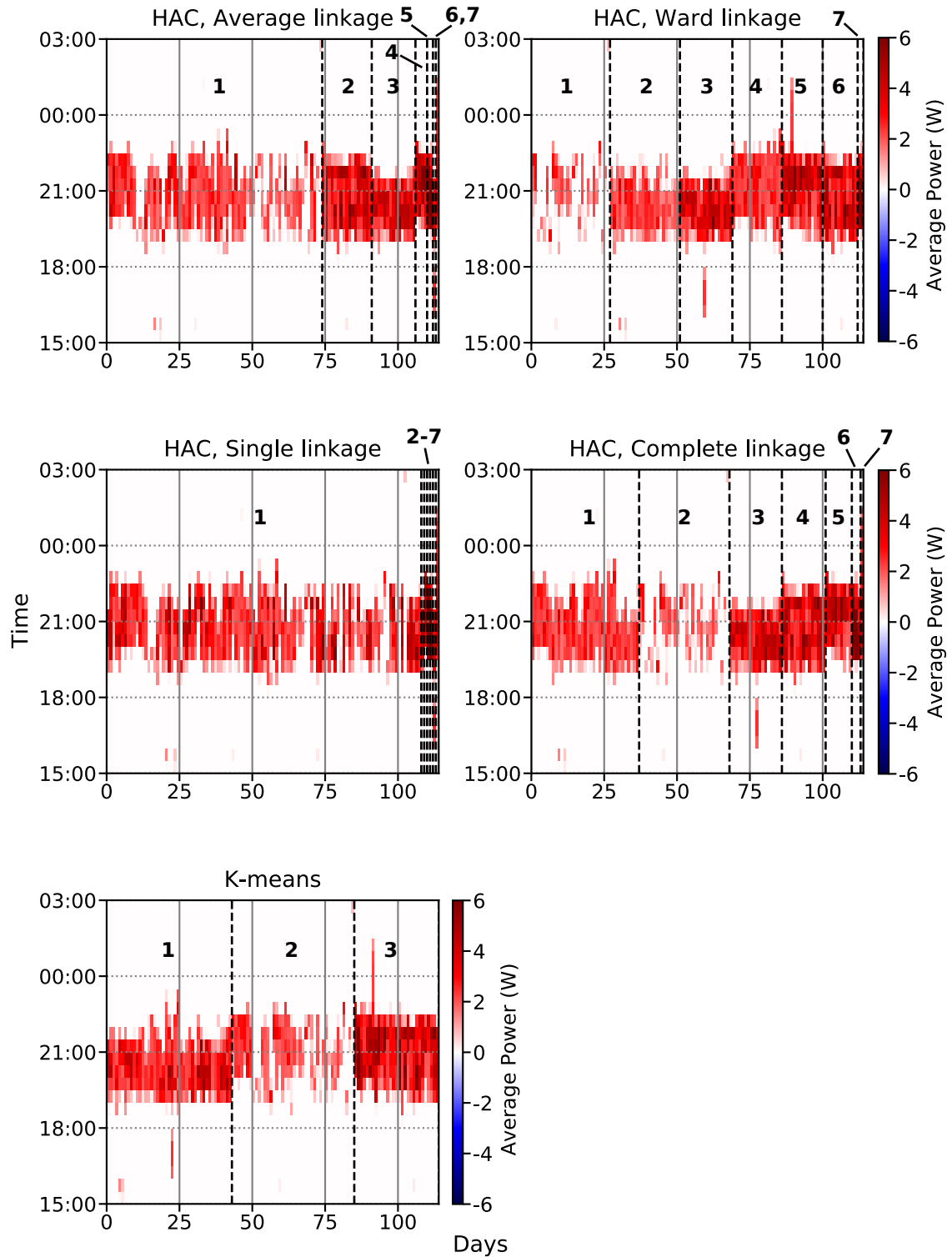


Figure 6.4: Heat maps of days of electricity use for household 50, for five clustering algorithms to demonstrate the different algorithms. The days are grouped by clusters, indicated by the dotted lines and labelled numerically. Comparison of HAC with average, Ward, single, and complete linkages, and k-means algorithm. For this clustering stage, applied to each household individually, a solution with a mixture of cluster sizes is meaningful as it isolates the most different electricity use days, and provides a grouping of the remainder (See Section 6.2.1.2).

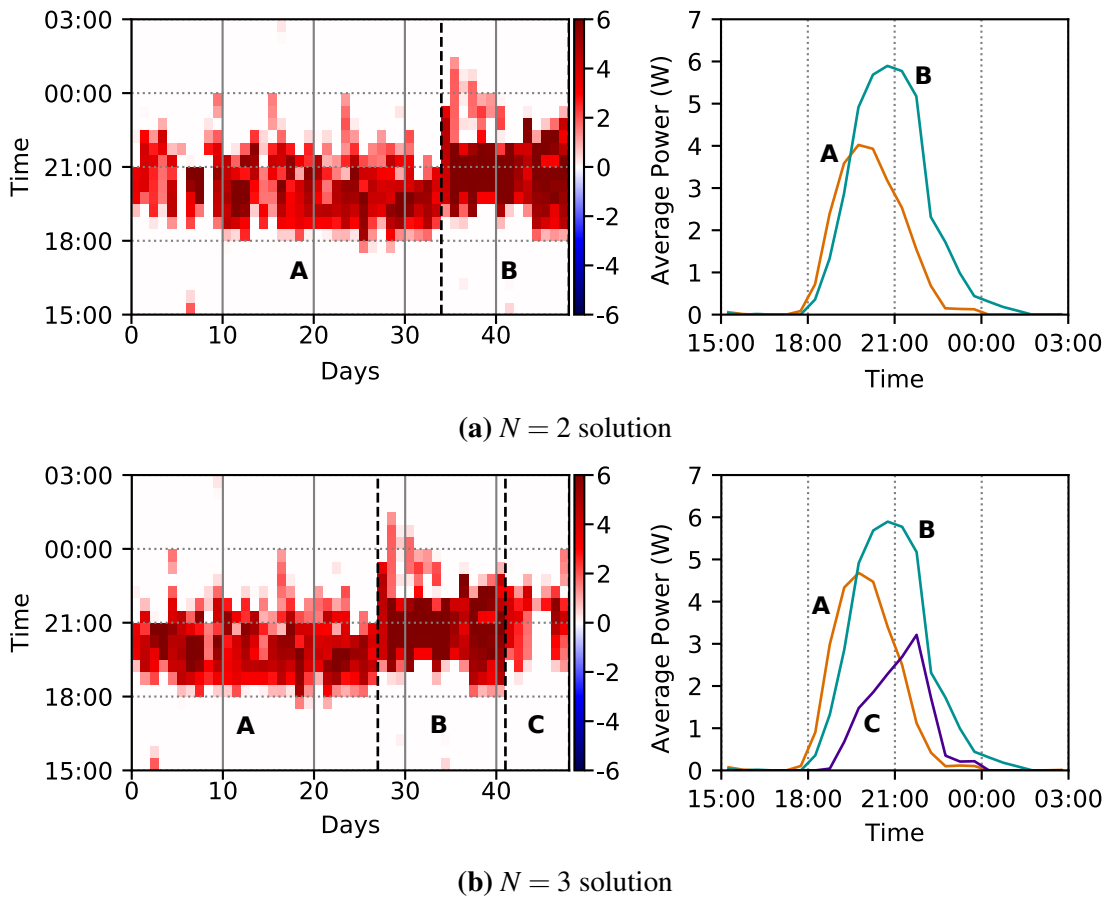


Figure 6.5: Heat maps and RLPs comparing results of clustering with HAC with average linkage for household 29. Results for a) $N = 2$, and b) $N = 3$. Heat maps sorted by clusters, labels to link RLPs with appropriate cluster on heat maps.

50 (Figure 6.6b), the largest cluster represents use patterns where less electricity is used compared to the second cluster.

The sizes of clusters found vary across the households, but in general there tends to be a larger cluster that represents the majority of days, along with smaller clusters that represent use patterns that are different from the norm.

The $N = 2$ solution resulted in 125 outlier days (5% of all use days across all households). Figure 6.7 shows a heat map of the outlier days, ordered by increasing daily electricity consumption. Noticeable types of outliers are visible on the heat map and are labelled, showing how differences in time of use and magnitude caused these days to come out as smaller clusters in the HAC process. On some days, use starts much earlier or much later than usual, and for a small number of days (6% of the outliers) there is lighting use during the afternoon. Some days showed higher than normal power use, whereas others had very low use. There are also examples of combinations of these factors.

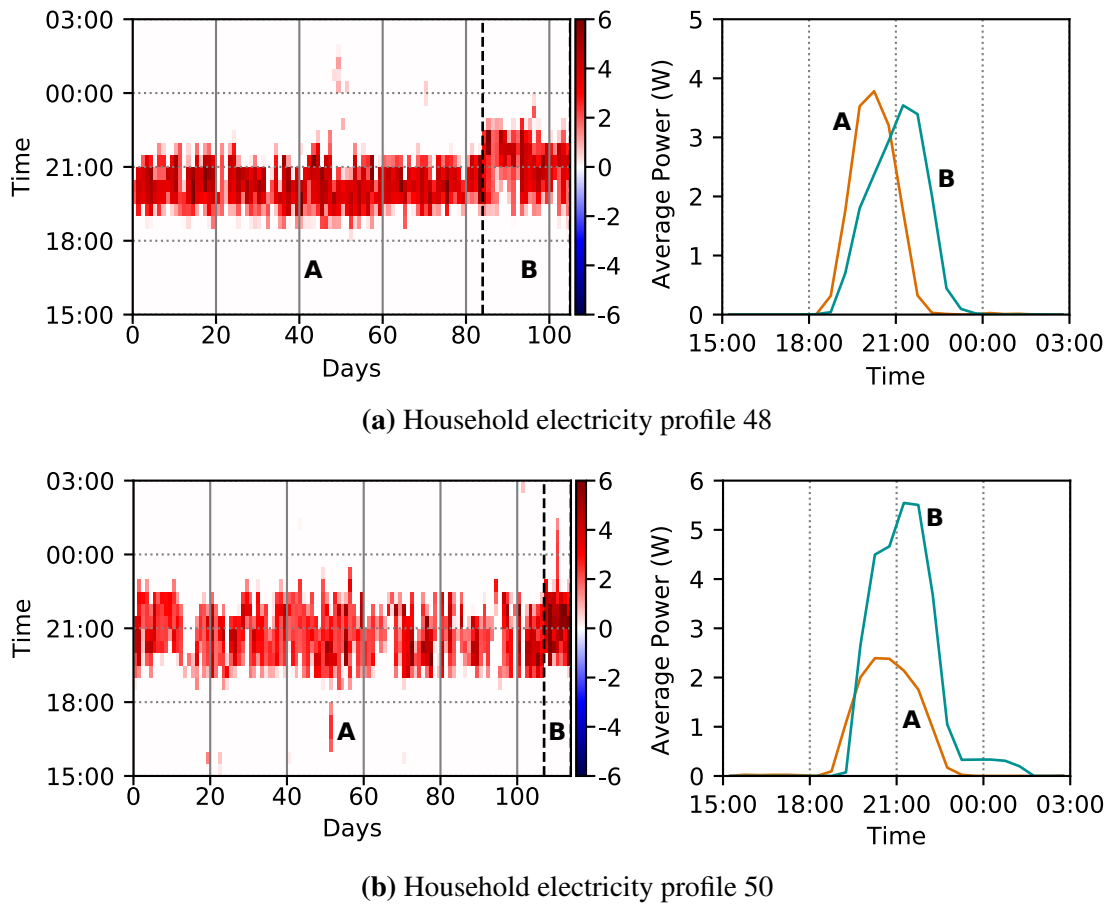


Figure 6.6: Heat maps and RLPs showing results of clustering with HAC with average linkage for two households. $N = 2$ result. Heat maps sorted by clusters, labels to link RLPs with appropriate cluster on heat maps.

Figure 6.8 shows the number of outlier days for each day of the week and month. Also marked is mean number of outliers across the categories of each plot, and the bounds of up to ± 2 standard deviations around the mean. Figure 6.8a shows that outliers were more common on a Saturday, as the number of outliers that occurred on a Saturday is approximately 1.75 standard deviations above the mean. This suggests that households deviated from their normal electricity behaviour on a Saturday more often than on other days.

Figure 6.8b shows that the majority of outliers occurred in July, and the number tails off over the months. Outliers may have occurred more frequently early on in the project as users tested the system before settling down into a more regular use pattern. But also, the number of batteries in use decreased throughout the monitoring period, as more were beset with technical problems.

To characterize the type of outliers further, and to see if certain types were linked with particular loading conditions, k-means clustering was used on the outliers. The $k = 2$ solution

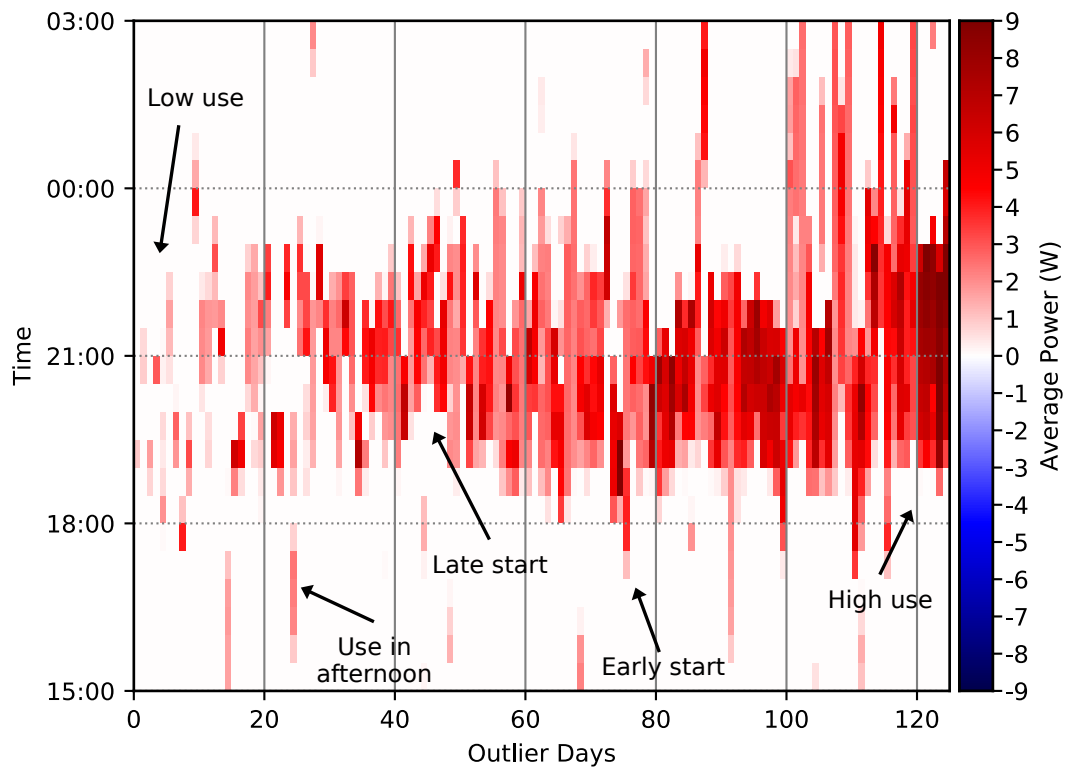


Figure 6.7: Heat maps showing the outlier days, with outliers ordered from smallest to largest daily electricity consumption.

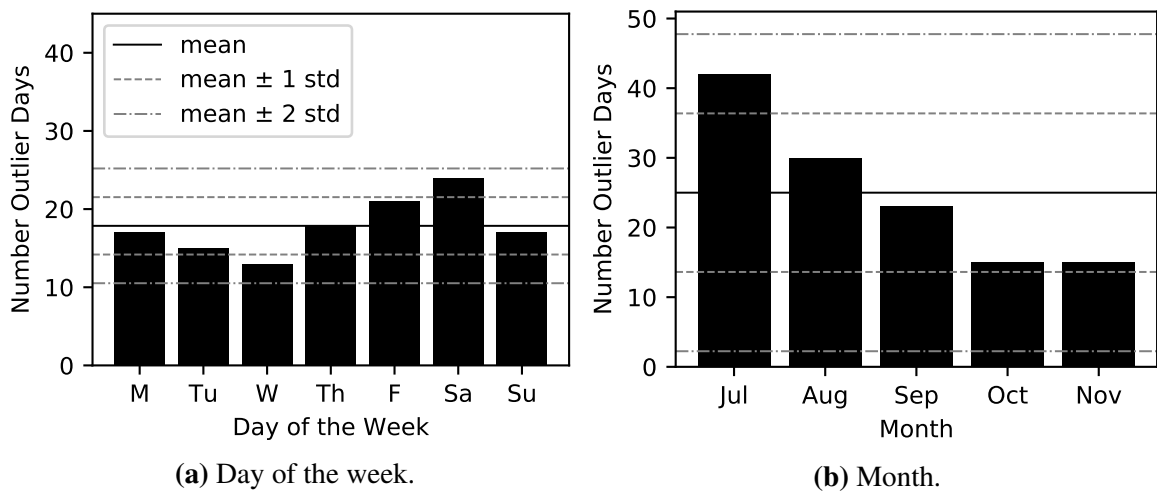


Figure 6.8: Bar charts showing the number of outlier days for a) each day of the week and b) each month. Lines show the mean number of outliers, and up to ± 2 standard deviations (std).

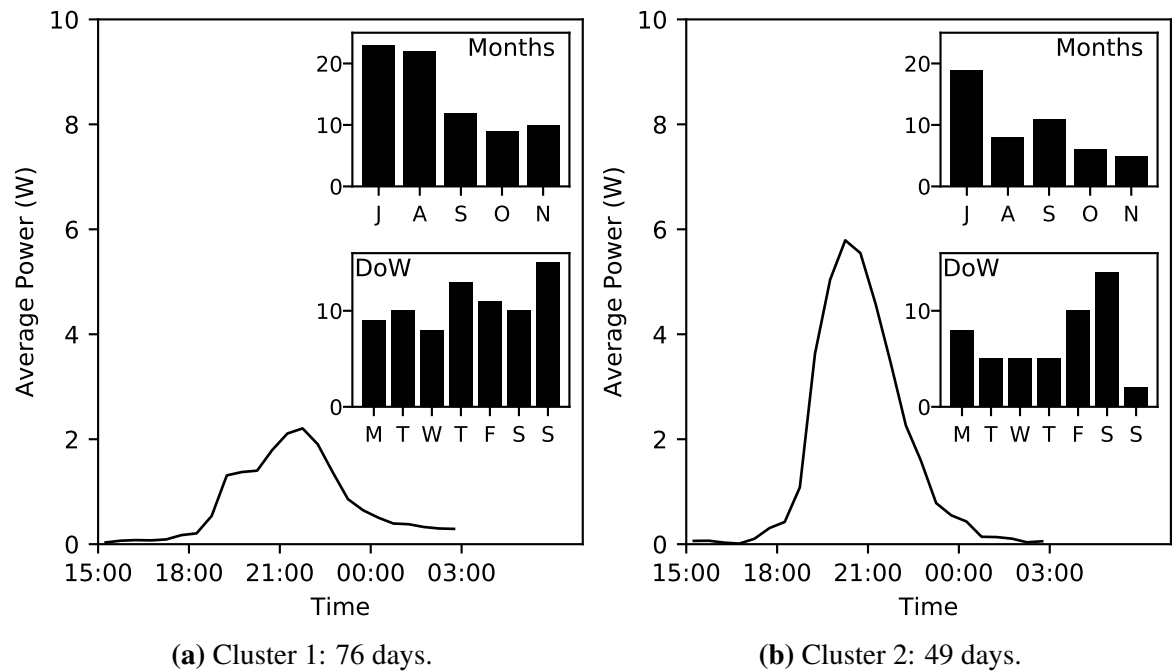


Figure 6.9: Centroids of the clusters formed when outliers are clustered using k-means, $k = 2$ solution. For each centroid, inset bar charts show the distribution of days according to day of the week (DoW) and months.

is shown in Figure 6.9, showing the centroids of each of the clusters formed (main plots). Magnitude, rather than time of use, has been the differentiator between clusters, as the outliers have been split into higher energy outliers and lower energy outliers. At higher k values clusters emerge that are distinguished by time of use, however for the sake of space these are not shown.

The inset plots show the distribution of day of the week and month for the sample of outlier days assigned to each cluster. Contrasting the day of the week bar charts shows that the most common day for high use outliers is Saturday, and only one high use outlier day occurred on a Sunday. Outliers occurring on a Sunday were almost always low use outliers. This adds to the previous insight, that suggested that household deviate from normal electricity use patterns on a Saturday more often than on other days, suggesting that when they deviate on a Saturday, they usually use more electricity than usual.

The outlier analysis has given some interesting social observations that point to differences in behaviour on the household level on a weekly basis. However, given the outliers are defined relative to each individual household, they represent the most different electricity use days relative to each household but not with respect to the whole community. Therefore,

we do not think these outliers as defined here have implications for the design of the technical electricity system.

6.3.2 Communal Energy Modes Across Households

The household RLPs were clustered together to find groups of similar RLPs across the household population. The aim of the analysis was to uncover communal energy modes that are common across the set of households. The input data set consists of 98 household RLPs.

6.3.2.1 Clustering Method Selection

The methodology in Section 6.2 was used to determine the most suitable clustering method and solution. Drawing on experience with the linkage criteria in Section 6.3.1, HAC was used only with the average linkage criterion, because the chaining behaviour of single linkage is not appropriate for forming groups, and complete and Ward's criteria gave comparable solutions to k-means. As before, a summary of the results is given along with justification of method selection.

Figure 6.10 shows the heat maps for the clustering solutions of the household RLPs for k-means $k = 4$, and HAC with average linkage for $k = 7$. The heat maps show that k-means algorithm gives more equally sized clusters compared to HAC with average linkage. The largest cluster consists of low magnitude household RLPs, the second largest consists of medium household RLPs, and the third largest cluster groups together the higher household RLPs. The fourth cluster contains household RLPs that start later in the evening and go on later into the night. By contrast, HAC with average linkage finds a larger collection of low household RLPs (65% of the total), and the remainder are in two reasonably sized groups or in single or two-member clusters. Given the purpose of this analysis is to uncover communal energy modes common across all users, identification of 'outlier' household RLPs is not relevant in this analysis. Therefore, the k-means algorithm was chosen to group the household RLPs together, to uncover communal energy modes.

To investigate the use of the elbow method to determine the 'best' number of clusters, Figure 6.11 shows a plot of the inertia of the k-means clustering solutions against number of clusters. In contrast to Figure 6.2, which demonstrates a strong elbow at $k = 3$ using mock data, the real data shows a vague elbow at $k = 2$, and an even less discernible pivot at $k = 4$. There is no strong elbow for this real data because there is not an underlying ground truth

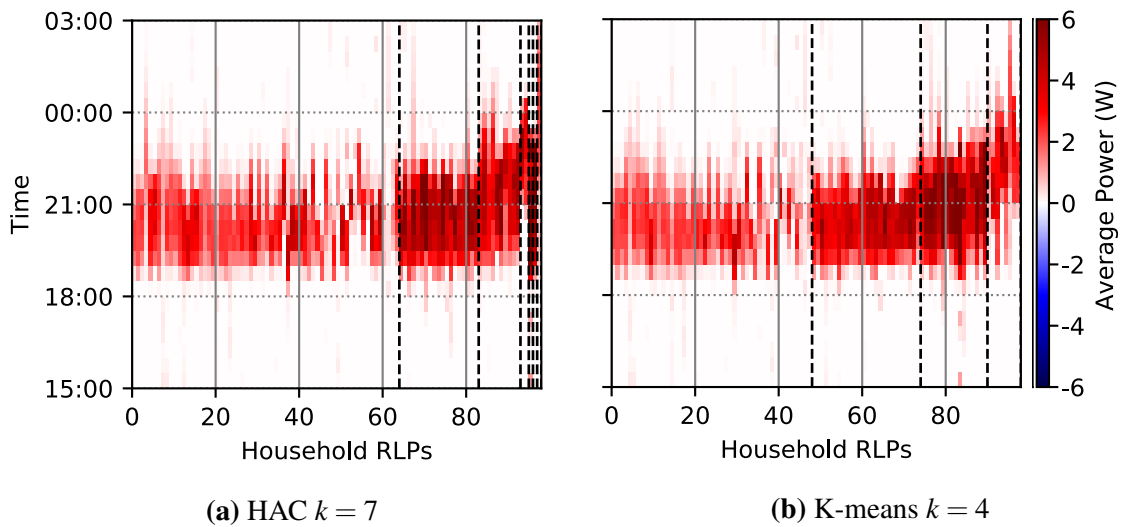


Figure 6.10: Heat maps comparing clustering algorithms on household RLPs: a) HAC with average linkage, b) k-means.

classification that the k-means algorithm can find (within the values of $k = 1 - 6$). This is not unexpected given the ways electricity can be used in a household are likely to be on a spectrum of different lengths of time and magnitudes, rather than in distinct modes of use that are very different to each other. As discussed in Chapter 2, the Literature Review, Section 2.3.1.4, and Section 6.2.2.3, visualization of the results is used in conjunction with the elbow method to discern a good clustering solution and is particularly important when there is no strong elbow. Inspection of the centroids and heat maps of the k-means solutions showed that $k = 4$ gave a good indication of diversity in the data set, giving a distinct set of centroids.

6.3.2.2 Results

Using the k-means algorithm, the household RLPs were clustered into four groups. Figure 6.12 shows the communal energy modes that are the centroids of the clusters formed. The modes are arranged in order of decreasing total electricity consumption.

The communal energy modes differ by time of electricity use and magnitude of electricity use, and the differences were used to label them. Mode 3 is distinguishable by the late time at which usage occurs relative to the others. Peak average power occurs at 22:00 and use extends to around 02:00–03:00, spanning approximately 8 hours. Labelled ‘late-peak’, this leaves three modes clearly distinguishable by the magnitude of the power values reached: high, medium, and low. Within these three modes, the time of use shows a trend. The time of peak power shifts later with increasing total energy magnitudes: low mode peaks at 19:30, medium

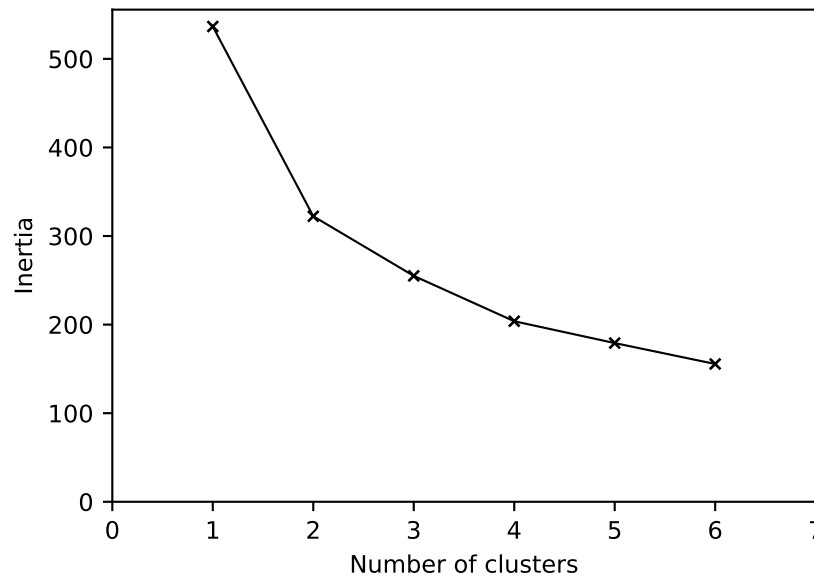


Figure 6.11: Plot of inertia of each clustering solution using the k-means algorithm against the number of clusters, to investigate using the elbow method to discern the ‘best’ number of clusters. A vague elbow is seen at $k = 2$.

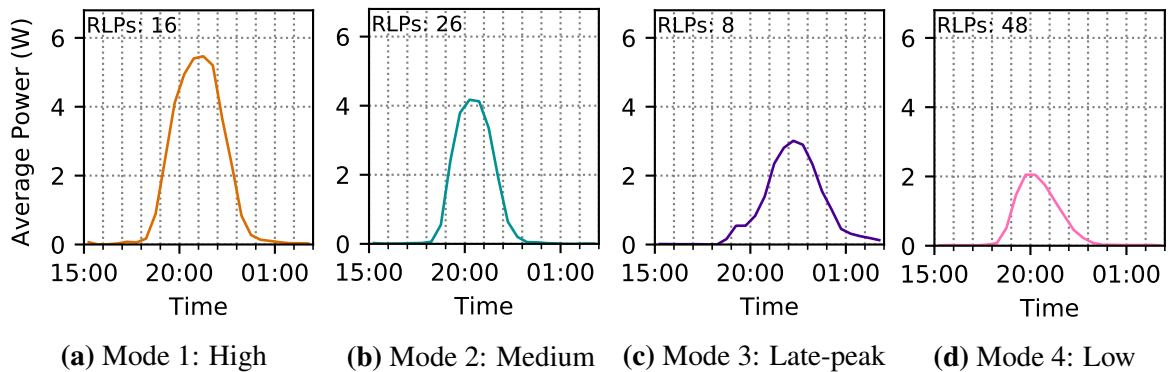


Figure 6.12: Communal energy modes resulting from clustering the household RLPs, with the number of household RLPs in each cluster noted. Ordered by total electricity consumption for each mode.

mode at 20:00, high mode at 21:00. The duration of use also varies. While low and medium modes have a similar duration of electricity use of 4.5 hours from around 18:30–23:00, for high mode electricity use extends to 00:00, a duration of 5.5 hours.

For comparison purposes, Figure 6.13 shows the average use pattern across all 49 households. It is calculated by first finding one RLP for each households’ electricity data by averaging over all the days associated with the household. The 49 average household RLPs are then averaged to give the a RLP that represents the average behaviour across the households.

The average household RLP across all households in Figure 6.13 has a peak average power of 3W, which is the same as late-peak mode, but the time of use is earlier in the evening, similar to low mode. It does not represent the higher magnitudes of energy use shown by

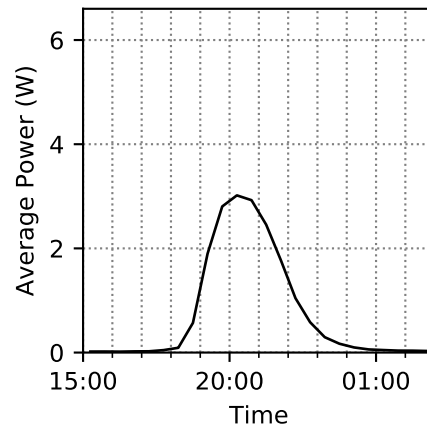


Figure 6.13: Average load profile calculated across all 49 households.

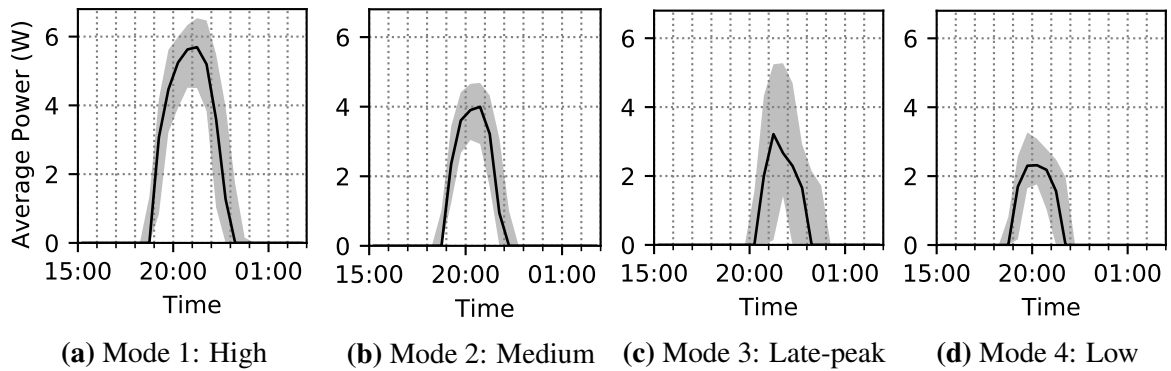


Figure 6.14: Variance in each communal energy mode: median (solid black line) and interquartile range (gray shading).

high and medium modes, the lower energy magnitude captured by low mode, the longer and later use duration captured by late-peak mode, or the changing time of peak average power draw shown by the range of communal modes.

The variance of the electricity use within each communal energy mode can be examined by tracing the days of use that are represented by the household RLPs assigned to each mode. This gives a collection of days for each energy mode. For each half-hour period, the median and quartiles can be calculated. The resulting plot for each mode is shown in Figure 6.14, where the solid black line represents the median and shaded section shows the interquartile range (IQR). Late-peak mode has more variance than the other modes, which is because it is less common, and so is represented by fewer days than the other modes. In general, the distributions of days represented by each mode are quite tight around the median, and for high mode the 75th percentile reaches 6.2W.

The two-stage clustering analysis enables investigation of the combinations of communal

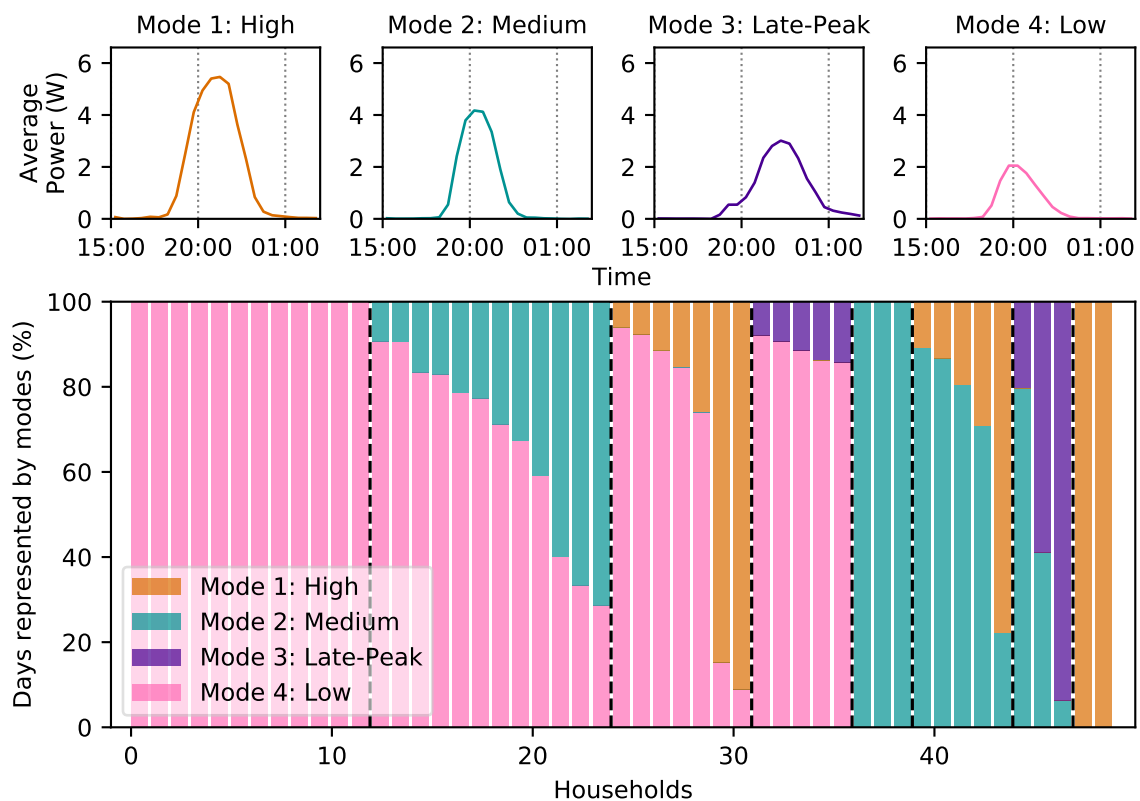


Figure 6.15: Bar chart showing the percentage of days, for each household, represented by each communal energy mode. Communal energy modes shown above. Bar chart segmented with dotted lines, marking groups of households with the same communal mode combinations.

energy modes used by households, and what proportion of time they spend using them. Figure 6.15 shows the distribution of communal energy modes used by the 49 households. Each bar corresponds to one household, and the color of the bar indicates the communal energy modes used in that household, with the proportion of the bar in each color reflecting the proportion of days that household spent using that mode. This was achieved by tracing the household RLPs to the communal energy modes they are represented by, and, for each household, examining the number of days represented by each household RLP. The bars are ordered by the size of the clusters forming each mode. Dotted lines segment groups of households with the same combinations of modes.

Table 6.1 accompanies the bar chart, detailing the number and percentages of households that use each communal energy mode (the first row), and the number and percentages of households that have each mode as their ‘dominant’ mode (the second row). The dominant mode for a household is the mode used for the highest proportion of time. The term ‘minority’ mode is used to describe modes used less frequently. The table provides the

Table 6.1: Number and percentage of households using each communal energy mode, and the percentage for which the mode is dominant

	Low	Medium	High	Late-peak
Households using the mode	36 73%	23 47%	14 29%	8 16%
Households for which the mode is dominant	31 63%	11 22%	5 10%	2 4%

numerical data behind the visualization of the communal energy mode distributions for each household in Figure 6.15.

The results show that some modes are used by more households, and for greater proportions of time, than others. Figure 6.15 shows that low mode, represented in pink, is the most popular mode in terms of the number of households that use it - 73% of households. Out of those that use it, the pink proportion of the bar is frequently larger than the other colors. This is because out of the 73% of households that use low mode to some extent, for the majority (86% of the sample, 63% of all households), low mode is their dominant use mode.

After low mode, medium mode is the next most common, and high and late-peak modes are the least common. When used, high and late-peak are also most frequently minority modes for households, shown by the small proportion of each household bar they take up on Figure 6.15. Late-peak is used by 16% of households, but is only the dominant use mode for 4% households. High mode is used by 29% households, but is the dominant mode for just 10% households.

The graphs also show the combinations of communal energy modes that exist and how prevalent each combination is across the data set. Some combinations are more common than others, suggesting that some patterns of behaviour occur together more frequently than others. The two most common combinations are low + medium modes and low mode only.

Some modes appear in all possible combinations, whereas some combinations do not occur at all. Low mode appears in combinations with each of the three other modes, and only two of 10 possible combinations do not occur: late-peak only and late-peak + high modes. These are less frequent modes that are usually the minority modes when they do occur.

It is also informative to consider the number of communal energy modes required to represent the RLPs of each household. 65% of the households were represented by two communal energy modes, and 35% by one mode.

6.4 Discussion

The first stage of the clustering analysis within household electricity data has shown that two or three household RLPs are sufficient to represent each household's electricity use (though only the results for $N = 2$ have been shown here). The second stage of the clustering analysis, clustering household RLPs, showed that four communal energy modes represent the use patterns across all the households. This is a small number of load profiles to represent residential electricity use, both in terms of individual households, and across the data set. Jin et al. [107] investigated various clustering methods on US residential data and found solutions with 20–90 clusters for k-means and HAC algorithms. In fact, in the literature, clustering of individual residential profiles is relatively rare, because they are often very volatile with large variations in power (on the kW scale). When clustering is performed on residential data, it is usually on aggregated residential profiles [96]. However, the household electricity data in this investigation comes from tier 1 users, where electricity is used for short durations and at low powers, so there is much less diversity in the use patterns than for urban/Global North residential profiles. Rather than using many high-powered appliances, which causes the volatility in individual residential profiles for middle and high-income households, these households are only powering lighting with the household battery 12V port, so variation in power use is low.

The results also show that there is more diversity across the set of households compared to within each household, because there are more communal energy modes than household RLPs for each household. This is to be expected; the range of behaviour within a household is likely to be smaller than across several households, as differences in practices and lifestyles will cause electricity to be used differently. This is also evident from the number of communal energy modes required to represent each household. For 35% of households, both household RLPs were clustered into the same communal energy mode.

The result that a relatively small number of load profiles are adequate to represent individual and grouped household electricity use is expected to extend to other electricity access systems beyond battery-charging systems. Households using solar home systems, or low power micro-grids, offering a similar level of electricity service provision, would likely have similar variation in electricity use. As households move up energy tiers, their electricity

patterns will become more diverse as they invest in higher power appliances. As such more household RLPs will be required to describe their different use patterns.

The different communal modes imply different practices are occurring in households. The shapes of the communal energy modes give insight into the resulting lighting use, relating to the number of lights in use and the time and duration of use. These have implications for the design of the battery-charging system and other electricity access systems.

The observation that all the communal modes show lighting electricity use starting at approximately 18:30, which coincides with sunset, has implications for electricity systems using solar irradiance to generate the electricity. There were very few days that showed electricity use during sunlight hours; almost all lighting is required when the sun is setting or has set. Therefore, for wired electricity systems such as micro-grids that are powered from solar energy, a storage component is crucial to bridge the time shift between generation and demand.

It was observed that the time of peak power use, and the time at which electricity use stops for the night, gets increasingly later moving from low, medium and high modes. Assuming that the end of lighting use in the evening occurs when the household is asleep, this indicates varying sleeping patterns of household occupants. The modes with lower total electricity consumption reflect evenings where the household goes to bed earlier than in evenings represented by higher modes, therefore using less electricity during the evening, earlier in the evening.

The magnitude of low mode, at 2W peak average power, is less than if one light was on for the half-hour period. This means that the days contributing to low mode do not have lighting electricity use for all the half-hour periods in the evening - on some days, the system isn't used at certain times. Low mode evidently gathers the days where lighting electricity use is low, but also not used consistently throughout the evening. We conjecture that this might occur if the battery lighting system is being used specifically for task lighting, so it is only turned on for portions of the evening. By comparison, medium and high mode represent days where more lights are in use.

It is notable that high mode does not correspond to all three lights being used at the same time for a significant proportion of the evening, as it does not reach 9W average power. This was an expected use of the system, drawing on data gathered in the community consultations (see Chapter 3), from which we understood that most households had two rooms, and that security lights were important. From this, the household lighting system was specified to

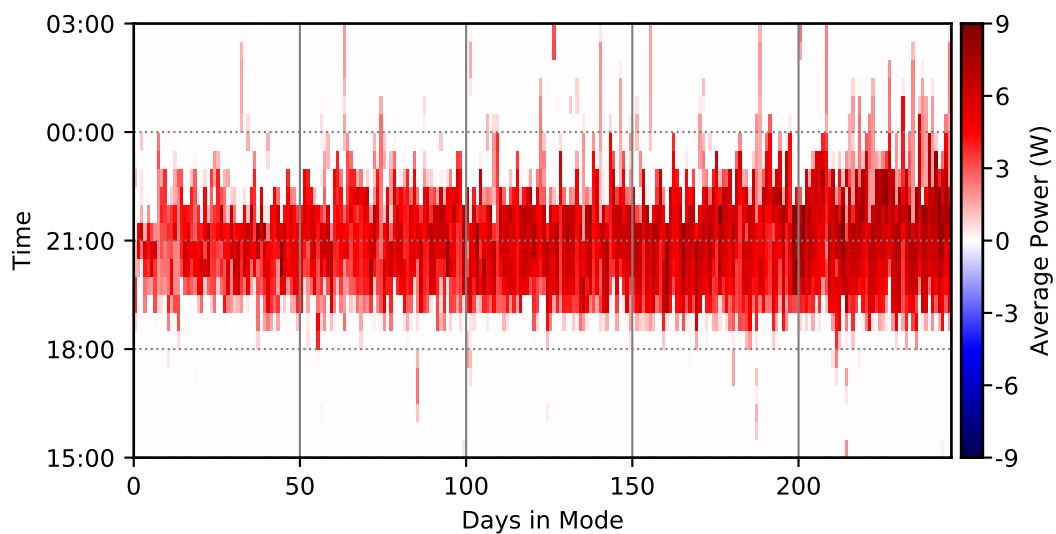


Figure 6.16: Heat map showing the raw days of data associated with high mode. The days are extracted from the household RLPs that were assigned to the high mode cluster. The days are ordered by total daily electricity consumption.

include three lights so that each space could be lit. To examine this further, Figure 6.16 shows the days that contribute to high mode. These are the raw days of electricity use, represented by the household RLPs that were assigned to the high mode cluster. The plot shows that few days have 9W lighting electricity use for the majority of the evening. Despite the expectations from the data gathered in the community consultation, in practice, powering three lights for a significant portion of the evening was clearly not an important requirement of the service.

There are several possible reasons to explain why what occurred in practice was not what was expected according to consultations with the communities. It could be that end users reduced their lighting use in order to make the battery last longer between charges, if they viewed visiting the charging hub as inconvenient. This means they adjusted their practices down from their aspirations due to how the system operated. Evidence for this was explored in Chapter 4, Figure 4.14, which found a borderline significant relationship between distance from the hub and average daily electricity use. Alternatively, or in addition, household practices may mean that more than one room of the house is rarely used at a time, so there is no need for all three lights.

Given the lighting set was rarely used to its full potential of powering three lights throughout the evening, it may be possible to reduce the size of the lighting service, offering two lights rather than three, and so reduce the cost of the system while still fulfilling the lighting needs of the end users. Before this judgment can be made, the data should be examined

at finer time resolution to determine whether or not three lights are used in households, and a judgment made as to whether there is a need for three lights. This area forms a line of study for further work that leads on from the research in this thesis (see Chapter 8, Section 8.2).

The distribution of communal energy modes was also explored for each household. Low mode was the most common mode and was often the dominant mode for households that used it. Therefore, for many households low mode reflects their standard electricity use pattern for an evening. This is in line with the observations in Chapter 5, where a positive skew was observed for most of the probability density functions of daily electricity use - higher energy magnitudes were increasingly infrequent. For households that usually use lower modes, their minority pattern or patterns may correspond to days where evening activities are different from most days, and for some reason they require more energy.

Late-peak and the high mode are most often minority modes, and we conjecture that they align with less common behaviour in the household, which is also why they are not seen in combination. Assuming again that lighting electricity use indicates when occupants are active in the evening, late-peak mode may correspond to occupants returning home later at night than usual, and the majority of household activity occurring later than usual. It also suggests that occupants continue to be active into the early hours of the following morning. We suggest that this may be related to the location of occupants work activity - for example, if they work outside of the village. Alternatively, light may be required into the night for other reasons, such as the presence of infants who require care during the night. For those two households for which late-peak is a dominant pattern, it suggests that the behaviour that causes late-peak mode is a standard occurrence.

Late-peak mode is important for system design because it suggests occupants are active for approximately 8 hours during the evening, in contrast to the other modes that suggest activity occurring for 4.5-5.5 hours. Given late-peak mode is used by just 16% households, one option in design would be to disregard this minority mode and economize on the design of the electricity system by assuming electricity is required for 5 hours only. On the contrary, a more conservative approach, to ensure the system was capable of meeting the extremes of demand, would be to take 8 hours as the expected duration of use each evening. When designing a system for higher power and electricity services, the differences in these scenarios would become more significant. In general, late-peak mode gives a useful indication of the longest duration of electricity use that can be expected for these households.

The second least common mode, high mode, is used by 30% households but only dominant for 10% (five households). However, for two out of those five households, it represents both the household RLPs of the household. We conjecture that in these cases there is likely a difference in the norms and associated practices in the households compared to the rest of the set. This might be a preference or need for higher lighting levels, or it being necessary to light more than one room, which may be caused by higher occupancy, the regular occurrence of more activities happening in parallel, or activities that require higher task lighting.

It could be tempting to economize on the design of an electricity system by neglecting the power requirements shown by high mode, reducing the specification of the equipment and so reducing costs. However, this would impact the electricity use of nearly 30% of households, for at least a portion of the time. In households that use high mode as a minority mode, it may reflect activities that although infrequent, are important to the quality of life and livelihoods of the households. In addition, as seen in Chapter 5, it is the households with higher numbers of occupants that would be most affected by neglecting to design for high mode. Neglecting it in the design would fail to meet the needs of a particular demographic of household.

Uncovering the underlying practices associated with the communal energy modes requires further investigation to gather data on where the lights were placed in the household, and what activities correspond to their use. This information would provide reasons for the different levels of lighting use, which could give a better underpinning for design decisions about the lighting system. A lack of time, resources and difficulty of accessing the remote sites, meant that this did not happen in this case study. The location of the lights was sought through the survey conducted in November 2016, but none of the surveys were filled in with sufficient detail. This topic is another area for further work, and is discussed in Chapter 8, Section 8.2.

6.5 Summary

The research question addressed in this chapter was: what are the typical patterns of electricity use within and across rural, low-income users?

The analysis found that there are four modes of electricity use that describe the distribution of use patterns within and across the households in the case study data set. The modes describe different magnitudes of lighting electricity use, and different times and duration of use, which have implications for the design of electricity systems. The methodology used to extract the

modes retains the spectrum of typical use patterns for each individual household. This means the analysis obtains the range of load profiles present in the data set, rather than the average, or most common behaviour. The results show where designing for an average use pattern (Figure 6.13) would negatively impact certain households for a proportion or all of the time. It would also underestimate the magnitude and duration of electricity use.

The extent to which the results and insights from the case study analysis are applicable to other electricity systems is discussed in Chapter 4, Section 4.5.

How can electricity demand data and household demographic data be used to estimate load profiles?

7

Estimation of Load Profiles for Electricity Access Systems

Contents

7.1	Introduction	167
7.2	Estimation of Load Profiles from Electricity Data and Demographic Data	169
7.2.1	Method	170
7.2.2	Results for Case Study	171
7.2.3	Discussion	177
7.3	Synthesis of Measured Electricity data and End-User Aspirations . .	181
7.3.1	Method	182
7.3.2	Results for Case Study	183
7.3.3	Discussion	186
7.4	Summary	189

7.1 Introduction

The previous chapter set out a method to extract a concise set of typical load profiles from time series electricity data, and found four communal energy modes that represent use patterns for households in the solar nano-grid case study. The distribution of communal energy modes described the modes used by each household and the proportion of days they were used for. In Chapter 5, a methodology was outlined and demonstrated to find demographic links to total daily electricity consumption, and use the results to estimate end-use electricity requirements for battery-charging systems. To appropriately size generation, storage, and

distribution components of micro-grids for new communities it is important to have an accurate depiction of the expected load profile [84]. Therefore, this chapter seeks to use demographic data to estimate load profiles for new electricity systems, in part following the same approach as in Chapter 5.

This chapter consists of two parts to address the estimation of load profiles for electricity access systems. The first part (Section 7.2) deals with estimating load profiles using measured electricity data and easy-to-measure demographic data, drawing on the results from Chapter 6 and the methodology of Chapter 5. The analysis in Chapter 6 retained the variation between days in load profiles used in individual households. Therefore, the electricity use of each household was described by at least one communal energy mode. This means it is not possible to use existing load profile classification models that assign customers to just one load profile, as reviewed in Section 2.2.2.3, Chapter 2. Instead, links to demographic data are explored through single explanatory variables, to retain the representation of the variation of electricity use patterns within households. Having linked electricity use to demographic data, we outline a demand estimation tool to estimate the load profiles for a group of households. The method is demonstrated using the lighting electricity data set from the solar nano-grid case study. The second part of the chapter (Section 7.3) addresses the issue that the direct transfer of data to estimate electricity demand (as shown in Chapter 5, and Section 7.2 of this chapter) estimates demand at one point of time, but does not inform about how demand might grow over time. Therefore, to estimate the expected future load profile and daily electricity consumption, we explore the synthesis of measured electricity data and data about end-user aspirations as gathered through the Service Value Method (SVM). This represents electricity use when households have access to more electrical services. The approach builds on the process of forming aspirational load profiles from energy-use surveys, as described in Section 2.2.1 in Chapter 2. This is also demonstrated using the solar nano-grid case study, using the lighting electricity profiles from Chapter 6, and the data about aspirational services presented in Chapter 3. While this analysis indicates how demand might change in terms of magnitude, it does not address the time scale of potential demand growth, which is out of scope of this thesis.

7.2 Estimation of Load Profiles from Electricity Data and Demographic Data

The first step to estimate load profiles for new communities is the analysis of electricity data to find load profiles for types of households characterized by demographic data. Two methods are used to do this. The first method is similar in process to Chapter 5, but the detail is different, and is given in full where there are differences. It involves finding the distributions of communal energy modes and load profiles for households with different attributes, building on the clustering analysis in Chapter 6. The second method uses the raw data of measured days of electricity use for each household (the output of Chapter 4) to form average load profiles for each type of household. The second approach is used as a means to validate the results obtained through the two-stage clustering process, and to explore the variance in the load profiles of the different household categories.

The variance in load profiles is of interest because an important design decision in terms of sizing electricity systems is what proportion of the expected demand must be met by the system. This defines the expected reliability of the system, with ‘reliability’ defined as in [39] to refer to the ability of a system to provide the desired load on the scale of a given time interval (in this analysis, for each half-hour). When examining the variance of demand by considering the quartiles and particular percentiles of a load profile, this design decision equates to deciding what quantile of demand the system must meet. We present the median, quartiles and 95th percentile of the load profile for each category of household to show this choice affects the load profile to be designed for.

The second step to estimate load profiles for a new community is a demand estimation tool. Load profiles are required to design wired electricity access systems, and the example is framed as estimating the load profiles for a feeder (distribution line) of a micro-grid that serves a group of dwellings. The tool uses the outputs of the data analysis step to estimate the expected load profiles for a feeder, using easy-to-measure demographic data for the households served. The various methods for the data analysis step, and demand estimation step, are described in Section 7.2.1.

In Section 7.2.2, the analysis and demand estimation tool are demonstrated on the solar nano-grid case study. The two methods of analysis to form average load profiles for each household category are compared in Section 7.2.3. Also explored are the implications of

the difference in electricity use across the household categories. Finally, the results of the raw data analysis approach enable a discussion on the variance in the average load profiles, and the effect of demand diversity.

7.2.1 Method

The proposed method consists of two parts:

1. Analysis of the household load profiles and demographic data to determine the average load profiles for households with different demographic characteristics;
 - (a) using distributions of communal energy modes
 - (b) from the raw data
2. Demand estimation for a group of households, based on the energy mode distributions for different household types and demographic data for the households.

The electricity data required for the analysis are the communal energy modes for a particular community, where for each household the distribution of energy modes is known. Chapter 6 sets out a methodology for obtaining communal energy modes for time series household electricity data. Also required is the raw data of daily electricity use.

Demographic data covering household attributes such as age, gender, and occupation should be gathered and processed as described in Chapter 5, Section 5.2, to define household attributes and attribute categories.

Figure 7.1 shows the steps in the analysis of the household energy mode distributions to form energy mode distributions for household attribute categories. To form the distribution of communal energy modes for a particular category, the household mode distributions in that category are averaged. To calculate the average load profile for the category, a proportional average of the relevant communal energy modes are calculated according to the energy mode distribution of the category.

To conduct the analysis from the raw electricity data, for each household category the days for the appropriate households are gathered and the mean, median, quartiles, the 95th percentile for each half-hour are calculated. Each household should be equally represented, contributing the same number of days to the sample.

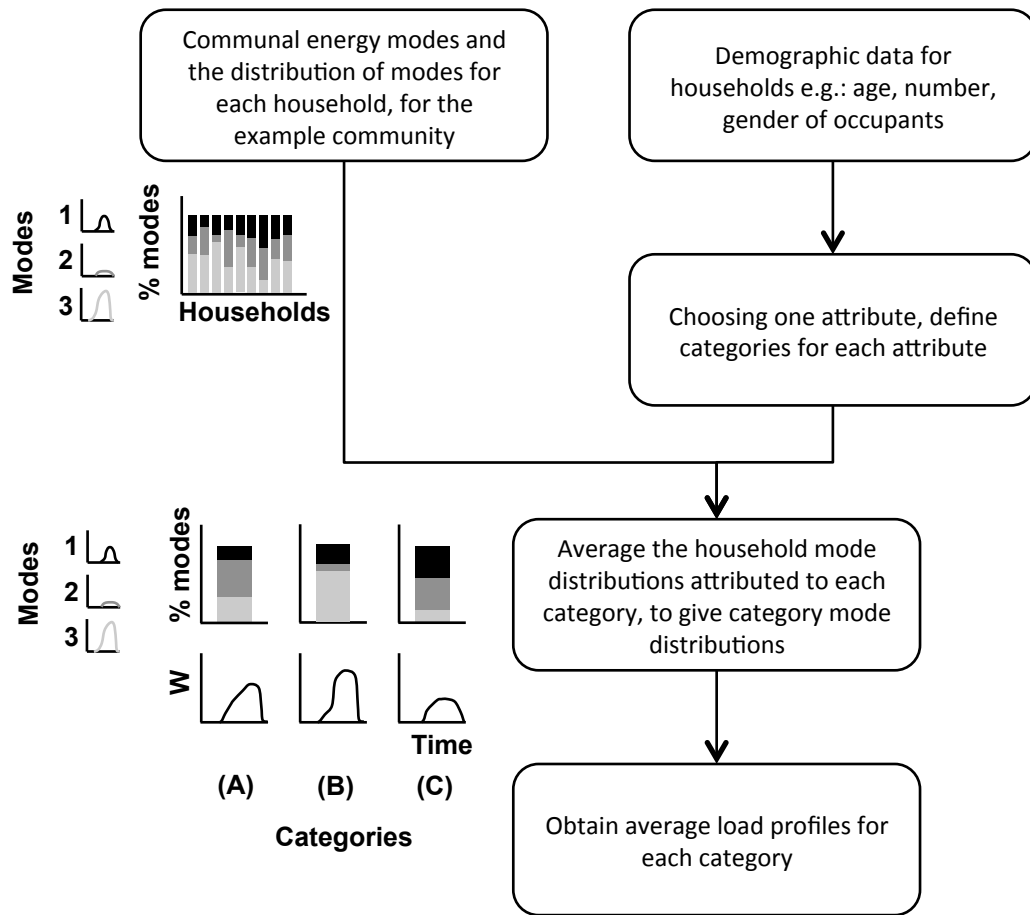


Figure 7.1: Method to analyze the household communal energy modes and demographic data to obtain the communal energy modes and average load profiles for different categories of households.

Figure 7.2 shows the steps in the demand estimation tool, using the category mode distributions to estimate the energy mode distribution drawn by the households on a feeder of a micro-grid. Also estimated is the average load profile for the feeder, which is the expected feeder load according to the distribution of energy modes.

7.2.2 Results for Case Study

The methodology to first analyze the electricity data and then estimate the distribution of communal energy modes and average load profile for a feeder of a micro-grid is demonstrated using data from the battery-charging systems in the thesis case study.

The acquisition of demographic data, and definition of attribute and attribute categories is described in Chapter 5, Section 5.3.1. The attribute used is the number of occupants in the household, and the four categories are: 1–4 people, 5–8 people, 9–12 people, Unknown.

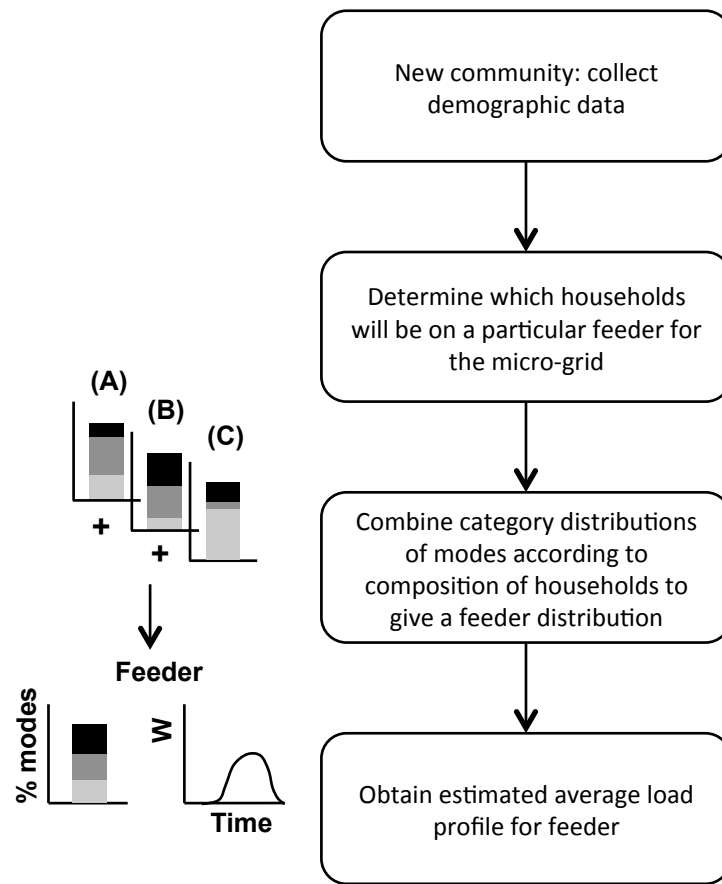


Figure 7.2: Demand estimation tool to estimate distribution of communal energy modes and average load profile for a micro-grid feeder.

7.2.2.1 Analysis: Demographic Data and Load Profiles from Communal Energy Modes

The methodology in Figure 7.1 was used to find the distribution of communal energy modes for the four number of occupants categories. Figure 7.3 shows these distributions for each household category, with the communal energy modes above for reference. The bars for each household category show the proportion of each mode (represented by different colors) used by households in that category.

The category energy mode distributions show that households with 9–12 occupants spend 60% of time in high mode, and otherwise use low mode. In contrast, households with 1–4 occupants spend 72% time in low mode, and most of the rest of the time in medium mode. Households with 5–8 people have a more even mode distribution, with just under 50% days in low mode, and the other 50% in high or medium modes. Late-peak mode makes a minor contribution to each category apart from 9–12 people, in which it does not appear. Its largest contribution is 7% of use in the 1–4 people category.

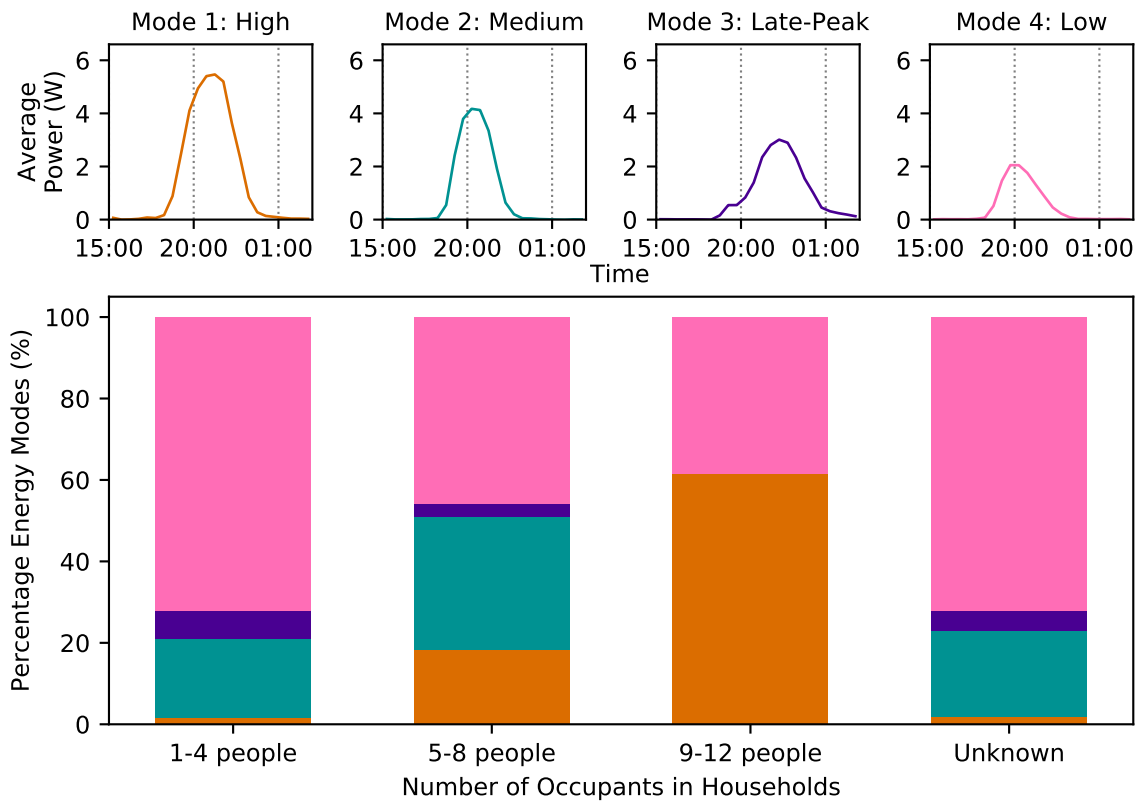


Figure 7.3: Distribution of communal energy modes for each household category. The communal energy modes are displayed above for reference.

Figure 7.4 shows the average load profile for a typical household of each category. As number of occupants increases, due to the increased proportion of high mode in the mode distributions, the load profile nearly doubles in magnitude, from 2.2W for households with 1–4 people (Figure 7.4a), to 4W for households with 9–12 people (Figure 7.4c). The time of peak shifts slightly later with increasing number of occupants but stays between 20:00 and 21:00. All the load profiles end around midnight, because late-peak mode is uncommon.

7.2.2.2 Analysis: Demographic Data and Load Profiles from Raw Data

For each category of number of occupants, the days for the appropriate households are gathered, but the challenge with this data set is that the number of use days varies across the households. Therefore, a sampling approach is used to ensure each household is equally represented. A fixed number of days from each household is sampled. This requires either sampling more days than exist for a household, or fewer days, depending on the number of active days of electricity use a household has (defined between 15:00–03:00 in this case).

If the number of active use days is n_d for a household, the number of days required was

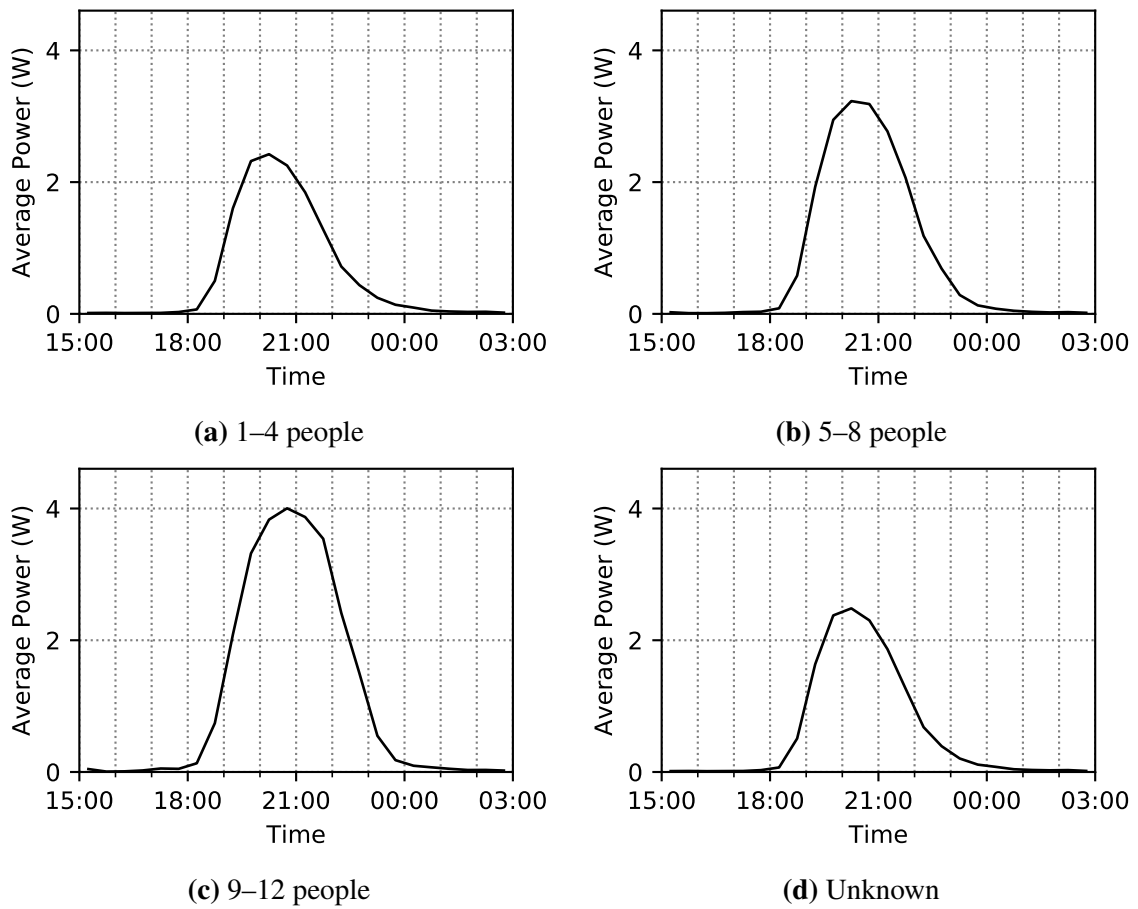


Figure 7.4: The average load profile for each number of occupants category, calculated from the distribution of communal energy modes.

chosen as the median of the number of active use days across the households, which is 46. Therefore, for the 26 households where $n_d \geq 46$ use days, 46 days were randomly sampled (without replacement). For the 23 households where $n_d < 46$, rounding down $46/n_d$ gives the number of times the full set of days for a household must be repeated. The modulus of $46/n_d$ then gives the number of days that must be randomly sampled (without replacement) from the full set of days, to make up the set of 46 days for that household.

For each category, the 46 days for the households in that category were combined. Figure 7.5 plots the mean, median, IQR, and the 95th percentile for each half-hour power value. Tracking the mean and median across the household categories, the expected trend of increased electricity use with higher occupancies is apparent. However, comparison of the mean load profile with those derived through the clustering approach in Figure 7.4 shows that the mean profile varies less across the categories than when calculated through the clustering process. For the 9–12 category, the mean profile does not reach 4W, peaking at 3.5W, and

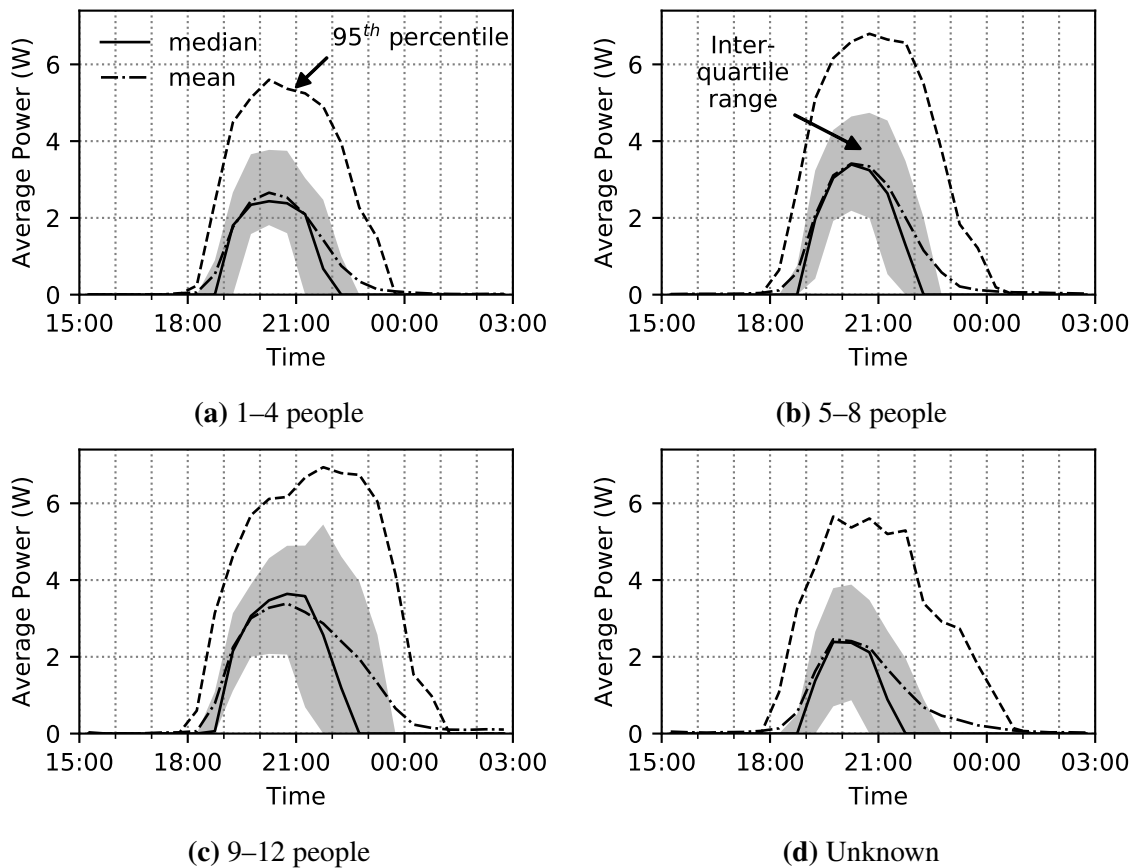


Figure 7.5: The load profiles and variance in load profiles for each household category, derived from raw data. The median (solid black line), interquartile range (shaded gray), 95th percentile (dashed black line). The days for each category consist of 46 days from each household in the category.

shows higher electricity use later in the evening. This is because of the over and under-sampling required to ensure each household contributes equally to the categories, and the small group size of the 9–12 category.

The 95th percentile also shows an increase in magnitude between households with 1–4 people and households with more occupants, but is approximately the same magnitude for 5–8 and 9–12 people households. However, the time of peak power use is later for the 9–12 people households, for all metrics.

The IQR, indicated by the shaded portion of each plot, shows where the central 50% of the use is in each category. The plots show that the variance in use for the 9–12 people households (Figure 7.5c) is more than for the two lower categories and the Unknown category, which is because of the small number of households in this category.

On top of this variation across categories, the IQR shows the same trend throughout the evening. The variance is lower at the start of the evening when households start using

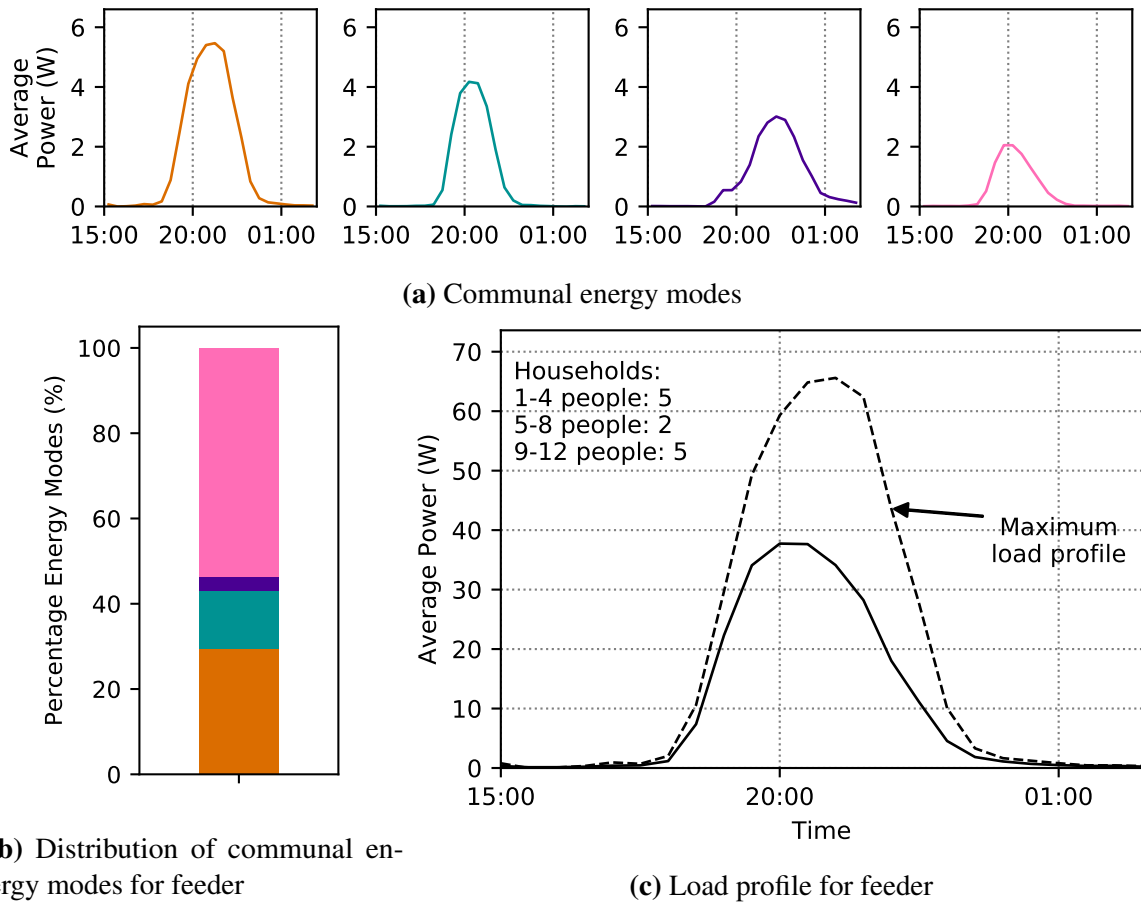


Figure 7.6: Result of demand estimation tool for a micro-grid feeder. a) Communal energy modes (shown for reference), b) bar chart showing distribution of communal energy modes, c) average load profile for the feeder (solid black line) and maximum average load profile (dashed line).

electricity, and increases around peak use for the remainder of the evening.

7.2.2.3 Demand Estimation: Micro-grid Feeder

Using the distributions of communal energy modes for each category in Section 7.2.2.1, the demand estimation tool in Figure 7.2 was used to estimate the energy mode distribution and expected load profile for a feeder of a micro-grid for Community X. Figure 7.6 shows the output for a feeder serving 12 households, where the number of occupants has been surveyed for each household. Figure 7.6b shows the estimated energy mode distribution for the households on the feeder. Figure 7.6c shows the average load profile for the feeder and includes the maximum average load profile which would occur if all households used high mode at the same time. The number of households in each number of occupants category is also shown on this plot. The communal energy modes are shown above the two other subplots for reference.

The households on the feeder are expected to draw 54% low mode, 30% high mode, with

most of the remainder in medium mode. The feeder load profile shows a peak average power draw of 38W over a half-hour period. The most conservative estimate for the feeder load profile occurs under the assumption that all households use high mode at the same time. This raises the peak average power draw to 66W over a half-hour period.

7.2.3 Discussion

Two approaches were used to examine the link between the categories of number of occupants and the average household load profile. By carrying through the results from the clustering process in Chapter 6, the distribution of communal energy modes was found for each household category and the average load profiles derived from them. In addition, the average load profiles were derived directly from the raw data.

Both methods provide useful insight and have advantages and disadvantages. Used together, the raw data approach validates the methodology from the clustering process, as it gives approximately the same results and certainly the same trends in magnitude of average power use across the household categories. The raw data approach is also useful for investigating the variance of electricity use within each household category, and is computationally less complex than the clustering process. However, it involves over and under-sampling the available measured days for each household to ensure equal contributions from each, which means some data are not utilized and other measured days are over-represented. The clustering process uses all the data for each household, and reveals much more information about the typical patterns of electricity use within and across the set of households. This shows how electricity use varies within homes, and provides insight into the different behavioural patterns, which is not possible through the raw data approach.

This discussion is based on the results from both methods. Firstly, the results linking electricity use and number of occupants are discussed. Secondly, the variance in household electricity use (drawing on the results from the raw data analysis) is explored, along with the effect of demand diversity on the resulting average load profile when households are aggregated.

7.2.3.1 Electricity Use and Number of Occupants

The energy mode distributions for each household category show that the proportion of time households spent using high mode increases from 2% to 62%, from low to high number of occupants categories. In turn, this leads to an almost two-fold increase in peak power for the

average load profiles calculated from the mode distributions. This is in line with the trends seen in Chapter 5, where households with more occupants more frequently used more electricity per day. As discussed in Chapter 5, this difference in electricity use is hypothesized to be due to the different norms and practices of large households in contrast to small households. All households had the same material culture in terms of the household battery and three light bulbs, and the majority had the same two-room dwellings. We conjecture that more occupants mean that more and varied activities are occurring in the household in parallel, causing lighting to be used in multiple rooms, and/or multiple lights used in the same room.

The two-fold difference in average power for lighting between households of different sizes is important for several reasons. Firstly, it is important for sizing equipment at the household level. For example, sizing embedded storage components in households in micro-grids requires knowledge of the magnitude and time of use in the individual household, which is demonstrated in this chapter.

Secondly, these results suggest that the number of occupants is linked to magnitude of electricity use, and so for new projects in new communities, the distribution of households in terms of household size may impact the total estimated demand. Working with an average power profile for the households in this study would curtail the use of the results for circumstances where the distribution of household sizes is different, such as for communities where larger households are common.

Thirdly, the variation in electricity use for lighting across the household categories gives an indication of variation in future demand for electricity when households have access to more services. We have hypothesized that the difference in electricity use across the categories is due to different norms and practices within the households. The causes are human-driven, rather than due to the electricity access technology, which is the same for all households. It is expected that the same differences in norms and practices will translate to different use of higher energy and power services.

It is not expected that the magnitude of the difference in electricity use across household categories will remain exactly the same at higher electricity access tiers; further research would be required to investigate this. It is possible that the use of different services will be correlated with number of occupants to different degrees. The variation in electricity use for lighting is indicative of the variation possible just due to different behaviours in households. Even if households had exactly the same additional appliances (such as TV, fridge, sewing

machine), we hypothesize there would still be a difference in use of the appliances across household occupancy categories. Such a situation with the same material culture in each household might occur with SHSs, where commonly the appliances are supplied with the generation and storage equipment. In a less controlled environment, where electricity systems can be used with products available on the market, we expect larger households to acquire more appliances than smaller households, and the difference in electricity consumption to be apparent for this reason too.

It was observed that late-peak mode is absent from the 9–12 occupants category, and is a minority ($\leq 7\%$) mode in other occupancy categories. However, results from Chapter 6 show that it was used by 16% of the households, and for two it was the dominant mode. The results in this chapter imply that use of late-peak mode is not connected to household occupancy, but occurs in households for a different reason. It is speculated that use of late-peak mode is dependent on aspects of households that are not correlated with the number of occupants: such as the location of work activities for the household occupants, or the presence of household members who require care during the night (e.g.: infants).

7.2.3.2 Variance and Diversity in Electricity Use

The variance in average load profiles across the different household categories showed that variance is smaller earlier in the evening, when households start using electricity, in contrast to the rest of the use profile. The time that households most frequently start using electricity is also the same across the different categories of households, between 18:00 and 19:00. This suggests that there is more consistency across households for when electricity will start being used, as opposed to how much electricity will be used and the duration of use. The precise timing is likely in response to sunset, after which lighting is required to see in the households. The reliable start time of lighting electricity use is a useful insight for micro-grid systems, as it means that there is a predictable start time that can be used in system scheduling for this service.

Compared to the variance in the communal energy modes shown in Figure 6.14 (Chapter 6), the household category load profiles in this chapter have more variance. This is to be expected, because it has been shown that the category profiles are combinations of the modes. In addition, the peak average power indicated for households with the highest average load profile (9–12 people), is two-thirds of the peak average power indicated by high mode. This

is the effect of demand diversity in a single household, which reduces the peak average power shown for individual modes to a lower peak when modes are aggregated together.

There is also diversity across households, which becomes relevant when estimating load profiles for groups of households, such as for a feeder on a micro-grid. Assuming the most conservative case, which is that high mode is used by all households at once, does not account for the diversity within and across households, which means actual peak power/energy use for a group of households is lower. This is evident from the demand estimation case study for a feeder of a micro-grid (Figure 7.6), where the peak average power of the average load profile is 58% of what it would be if all households used high mode at the same time. Taking this conservative approach and neglecting diversity would lead to over-specification of an electricity system.

The diversity also impacts what quantile of electricity demand is set as the target load profile that the design must meet. We display the 95th percentile for the electricity use of each household category. However, the 95th percentile is defined according to each half-hour period, so does not represent the worst-case load profile on an actual day, and it is unlikely that a load profile of this scale would actually occur for a household on one day. Instead, a value on the 95th might be reached for one half-hour period, and then electricity use return to more normal levels. When aggregating multiple households, it becomes progressively less likely with increasing group size that the 95th percentile of average power use occurs simultaneously across households in reality. Therefore, diversity in use may mean that the 75th percentile is more appropriate as a design point, particularly for aggregated households as for a micro-grid.

There are several characteristics of electricity access projects in the Global South that may support this lower design point. Projects are resource constrained due to the low-income of users, and the need to be financially viable for developers. In addition, it may be possible to compromise on service reliability to reduce the cost of the service. The reliability of electricity provision from the grid in Kenya is not high, with frequent blackouts requiring customers to have backup generation [145]. This impacts the expectations of consumers, and they may be prepared for, and so more accepting of, a service that does not function 100% of a time, in contrast to electricity consumers in the Global North. Trading off reliability to increase affordability is a design approach advocated in [146], where a micro-grid sizing strategy includes analysis of cost versus reliability over a year and at sub-daily time resolutions. This enables designers with knowledge of local energy needs to make sizing

decisions based on an acceptable reliability reduction given a reduction in the cost of the system, which may make the system affordable for consumers. Therefore, if there were severe financial constraints in a project, setting the reliability design point lower may be the best way forward so as to make the project viable.

In contrast, in some projects it may be important to take account of expected growth in demand. In this situation, the design point may be placed at the 95th percentile and an additional factor of safety applied. It might be argued that every project in rural, off-grid scenarios in the Global South should be designed to expect demand growth, because it appears logical that electricity demand will increase once access is enabled. However, low socioeconomic circumstance of users can mean that they are unable to take advantage of higher levels of electricity access. Users do not necessarily experience demand growth unless their economic circumstances also change and they can afford it. There is also dependence on the supply chain of appliances, in terms of availability and affordability. This is partly why there is significant emphasis on enabling productive use services that can be used to generate income in electricity access projects [7]. It is also why there is a large emphasis on developing business models and financial mechanisms for electricity access systems to make them affordable for the end user [147]. Furthermore, experience with the SONG project has shown that the range of electricity access solutions, including grid extension and off-grid strategies, are not mutually exclusive, particularly when the grid is not a reliable source of electricity. Therefore, it may be possible to know, or plan, that a particular project is not expected to handle growth in demand, depending on the project context and how electrification is developing in the area.

7.3 Synthesis of Measured Electricity data and End-User Aspirations

The second part of this chapter draws together the measured electricity data, and the data collected about end-user needs, aspirations, and context, through the Service Value Method described in Chapter 3. The data gathered include quantitative data about the relative importance of household services to the end users, and qualitative data about aspirational practices and associated context. An aspirational load profile is constructed that indicates possible future demand if end users are able to access aspirational services and have the required electricity access to support them. The effect on the total daily electricity consumption is

explored, drawing on the analysis in Chapter 5, and the impact on the sizing of household batteries in battery-charging systems is discussed.

7.3.1 Method

The data gathered for the case study demonstrating the Service Value Method are used to choose aspirational household services. The household services that are in the easiest two Ease of Delivery categories (4 and 5) are selected from the Service Map shown in Figure 3.4 from Chapter 3, Section 3.3.2.

Although the electricity data set in this study is for lighting only, the timing and duration of use indicates when occupants are active in the households during the evenings. The method is based on the electricity use for lighting of a typical household in one of the household categories. This, along with the qualitative data from Chapter 3 (Appendix A), is used to construct a use case describing a typical day in an example household. The number of each appliance present in the home, placement in the home, and associated use patterns are derived from the use case. The load profile is constructed by combining the power ratings of the appliances when they are being used. The result is a load profile for a particular day; occasional appliance use is not accounted for.

For the case study, the load profile is constructed from half-hourly average power use readings. Duty factors are used to convert rated power values to half-hourly average power values for services that are controlled by a thermostat. This means that the load profile does not show instantaneous power, and so cannot be used to extract peak instantaneous power, which is required for sizing power electronic converters because of the short thermal time constant of these devices. For this case study, the focus is on the effect on the size of the storage components of battery-charging systems, and specifically, on the size of the household battery.

To explore the effect of the additional services on the sizing of a battery-charging system, the daily electricity consumption of each service are summed to give a total daily electricity requirement. This is added to the end-use electricity requirement for lighting for the corresponding household category in Chapter 5. Along with daily end-use electricity requirement, we explore the electricity required during the night, defined as a 12 hour period between 18:30–06:30. This indicates the magnitude of energy required from a battery, assuming that day-time electricity was provided via solar PV generation.

In this process, the effect of base loads is analyzed separately from services used actively. The additional daily electricity consumption is calculated twice, once without inclusion of the base loads, and once with them. This is because the interrupted power flow when the household battery requires charging at the hub makes it less feasible or sensible that a battery-charging system be used for household base loads (particularly for services which cannot be interrupted without damaging consequences, such as the incubator). Therefore, both scenarios are considered.

7.3.2 Results for Case Study

The use case is presented below. It is assumed occupants are active between 18:30–22:30, because the use case is about a family of four, and a lighting electricity profile for a household with 1–4 people is used. The heat maps presented in Appendix C show that the household battery was rarely used in the mornings. Therefore, electricity for lighting and other occupancy-dependent loads are ignored for the morning, but base loads are obviously present throughout the 24 hours.

Use Case *There is a young family of four of two children and their parents. They have been out all day: the children have been at school, the father to the market to sell produce, the mother working on the family's farm. The children and mother arrive home at 18:30. They switch the TV and fan on in the living room to have on in the background, and plug in the rechargeable torch used to visit the outside kitchen and latrine at night. A small fridge is used to store perishable food items. The family also have an incubator, which is currently incubating 12 eggs from the chickens in their yard.*

The mother spends half an hour mending the torn clothes of the children with the sewing machine while the children do their homework. At 19:00, the mother uses the shaver to shave her children's heads. She then prepares the evening meal from a separate building in the compound using a charcoal jiko (cook-stove), taking the torch with her as the jiko is placed outside. At 19:30, the father returns home, puts his mobile phone to charge, and puts a DVD on. He watches with the children while surfing the internet on the computer for news and information about market for their produce. At 21:00 the family eat while watching the end of the film, and the children are put to bed when it finishes. At 22:30 the parents go to bed, and the mother plugs her phone in overnight.

Table 7.1: Aspirational services from Kenya SVT data, and power, time of use, and placement or use details. Active occupancy: 18:30–22:30.

Appliance	Average Power (W)	Time	Placement or Use
Mobile phone charging	5	19:30–01:30	Two phones charged
TVs	8	Active occupancy	1 in living room
Torch	5	18:30–19:30, 21:00–22:30	Charging
Fridge	45 (0.3)	Constant	
Incubator	80 (0.66)	Constant	
Computer	27	19:30–21:00	1 in living room
DVD player	10	19:30–21:30	1 film watched
Fan	8	Active occupancy	1 in living room
Shaver	10	19:00–19:30	Personal use
Sewing machine	30	18:30–19:00	Personal use

The aspirational services drawn from the Service Map (Figure 3.4) are listed in Table 7.1 in order of their popularity, which in this case is defined as the number of counters each service received during the SVT. The radio is omitted because it does not feature in the use case (a radio is unlikely to be on at the same time as a TV in the same room). The table also outlines power ratings (with sources in Table A.3), the time the appliance is used for, and notes on placement or how it is used. For the fridge and incubator, factors that take account of the duty cycle of the appliances are included in the power column. The duty factor for the incubator was measured for an incubator with the same rated power. The duty factor for the fridge is from measured data in the literature [148].

Figure 7.7 shows the estimated load profile as a bar chart focused on the time span 15:00–03:00. Services are stacked vertically and marked out by color. Half of the power draw is from the base loads (fridge and incubator), and the remainder is due to active use of services by occupants.

Table 7.2 gives the daily electricity consumption for each service over a 24 hour period, and the total additional daily consumption (not including lighting). The calculation is performed twice, firstly neglecting base loads, and secondly including them. When base loads are included, we also show the total electricity consumption for a night, defined by a 12 hour span. The electricity consumption is dominated by the base loads for all cases.

Table 7.3 shows the end-use electricity requirement for just a lighting service (carried forward from Chapter 5 analysis) and for when the additional services are included, with and without inclusion of the base loads, for a day (24 hours) and a night (12 hours). Without

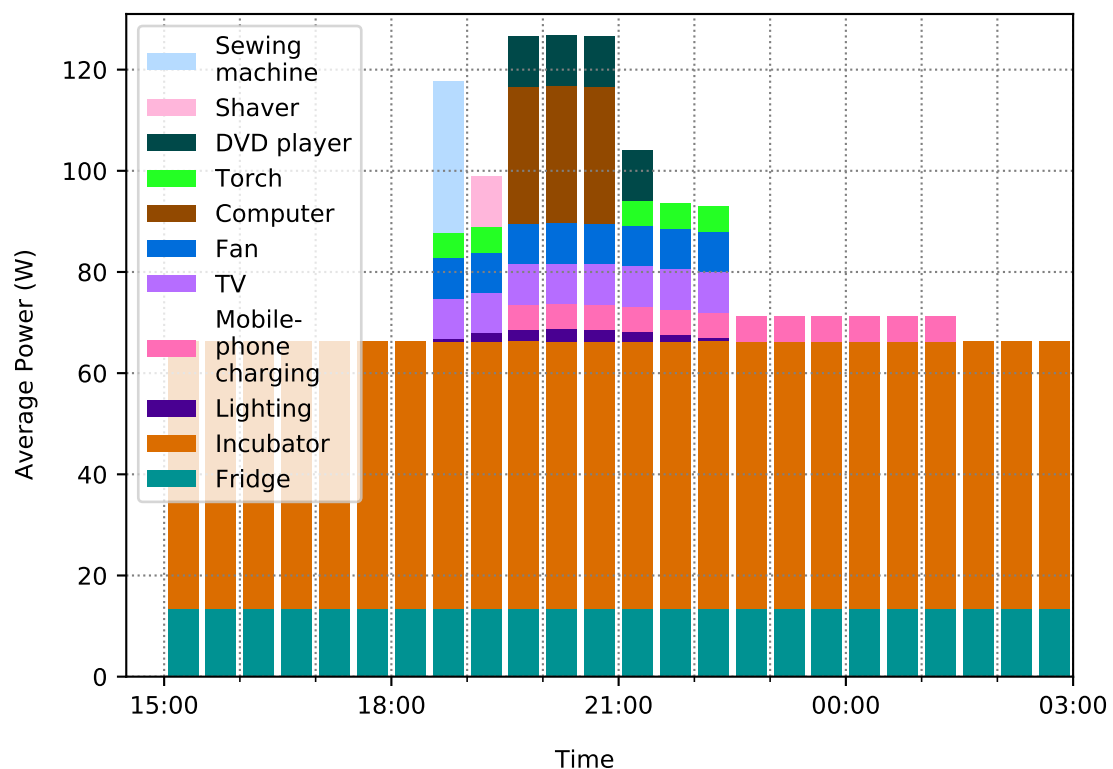


Figure 7.7: Estimated load profile combining measured electricity data and data about aspirations for appliances. Calculated for a household with 1–4 occupants with access to the services indicated important and feasible to deliver during the SVT process in Chapter 3. Shown for time span 15:00–03:00.

Table 7.2: Daily electricity consumption of additional aspirational services.

Appliance	daily electricity (Wh)
Mobile phone charging	30
TV	32
Torch	12.5
Computer	40.5
DVD player	20
Fan	32
Shaver	5
Sewing machine	15
Total	187
<i>Including base loads</i>	
Fridge	324
Incubator	1267.2
Daily total including base loads	1778.2
Night-time total including base loads	982.6

Table 7.3: Estimate of daily end-use electricity requirement for lighting and additional services.

	End-use electricity requirement (Wh/day per household)	
	Daily (24 hours)	Night (12 hours)
Lighting only		16
No base loads		203
Including base loads	1794.2	998.2

the base loads, the daily household end-use electricity requirement is approximately 200 Wh, which is nearly 13 times the lighting load alone. Including the base loads increases the electricity requirement to 1 kWh for the night, and 1.8 kWh for a whole day.

7.3.3 Discussion

The additional services, even without the base loads, represent a significant increase in daily electricity consumption. The scenario of electricity use that neglects the base loads is only just technically feasible for a battery-charging system incorporating flooded lead-acid batteries that are commonly used in rural Global South contexts. Such batteries observed in use in the communities with the solar nano-grids ranged in size between 35–65 Ah. A 12V 65 Ah battery is 780 Wh, which, using a 50% depth of discharge factor in order to not degrade the battery, and charged every two days, almost supports the use case scenario without the base loads (providing 95% of the daily electricity requirement each day). Therefore, the use case scenario without base loads is just physically feasible using large flooded lead-acid batteries, if end users charged the battery every other day. Improved lead-acid battery technologies such as Absorbant Glass Mat lead-acid or gel lead-acid batteries would enable a deeper depth of discharge to be obtained with reduced resulting degradation, and so could extend the time between charges to three days. However, these are more expensive and none were observed in visits to case study communities.

In addition, in reality, lead-acid batteries around 65 AH in capacity are heavy to lift, and although the extra service provision may act as an incentive for frequent charging, feedback from the end users in this study showed charging was found to be inconvenient. As discussed in Section 5.4 of Chapter 5, it is expected that end users will trade-off the utility of electricity access and the services enabled with the costs (financial and otherwise) of charging, and find an operating point that is appropriate for them. We hypothesize that end users would

actually charge less frequently than every two/three days, and would reduce their use of services from the use case description to make this work.

Furthermore, if able to afford the additional services, a household may also be able to afford to purchase a solar PV panel and construct a SHS to charge in situ, rather than using charging facilities at a local hub. Another piece of common feedback about the battery-charging hub in the solar nano-grids case study was that it would be better if the project could provide solar panels directly to the households, so end users did not have to visit the hub. It was not financially possible for the end users in the case study to purchase panels, but it is hypothesized that there is a stage of socioeconomic development that coincides with voluntary independence from a battery-charging system by purchasing solar panels. Once it becomes financially possible for end users to circumvent the inconvenience of taking a battery to charge at the hub by buying a solar panel themselves, they are likely to do so. This reinforces that for practical usage reasons, battery-charging systems are suitable as an early progression of electricity access, reaching tiers 1 or 2 on the Multi-tier Framework for electricity access (see Figure 1.1 in Chapter 1), but do not provide a permanent solution.

Alternatively, battery-charging systems coupled with centralized productive use services, as in the solar nano-grid model, may be more acceptable to end users and so longer-lasting. An equally important part of the solar nano-grid model was the presence of productive use services at the hub, alongside the battery-charging systems. As described in Chapter 3, some of the appliances installed at the hub were not suitable, and, partly due to disruption caused by insolvency of the NGO partner, it was several months before the project was able to address the issues affecting those services. Therefore, to the end users, the rationale of centralizing the electricity generation, to enable both the meeting of household electricity needs via a battery-charging hub, and provision of income-generation opportunities, was not evident in practice. It is therefore not surprising that there was feedback asking for the distribution of individual solar PV panels. If the productive use services had been active and successful from the start of the project, we hypothesize that the system set-up would have made more sense to end users, as they would have had the opportunity to benefit from the income generating opportunities.

As well as encouraging social acceptance of battery-charging systems, the presence of the central solar hub with working productive use services is also key to the justification of battery-charging systems over the alternative of providing many individual SHSs. Without the ‘productive use hub’, and given end-user preferences for in-household charging, it could

be argued that a better solution is to provide small solar PV panels alongside household batteries. There would also be an economic case for this if the combined cost of the generation components, power electronics, and distribution equipment, were equivalent or cheaper for the distributed SHSs scheme. This might be the case given the size of PV panel required at each household could be relatively small, as the whole usable battery capacity may not be required to be replenished each day. However, there is the potential for economies of scale in terms of charging equipment for the centralized scheme. A comparative cost analysis to determine the cheaper solution, which would require modelling the operation of each scheme (including optimization of the charging strategies), and sizing components to meet a given demand, is out of scope of this research.

In any case, given the presence of a functioning productive use hub, a battery-charging system located at the hub makes sound sense as it can utilize excess electricity generated at the hub during times of low demand for the productive use services. There is therefore the potential for a cost-saving on the generation infrastructure, compared to a scheme with a productive use hub and separate household electrification by SHSs, with PV panels at the hub and at the individual households.

The end-use electricity requirement including base loads, for a night or a whole day, gives an electricity consumption that is not manageable by the lead-acid batteries found in the case study communities. It is clear that several of these batteries would be required to power these base loads alongside the other services, which would require daily charging and careful scheduling to ensure no disruption. The time and effort required means this is not a practical solution, and a higher tier electricity system, such as a micro-grid, would be required to meet the load profile with base loads included.

This suggests that once households want to power base loads like fridges, a micro-grid is the next logical step leading from a battery-charging system. Battery-charging systems are better understood as temporary electricity access solutions, which serve a purpose of improving low-tier electricity access in households. Removing the cost of having to take the battery to charge each day, but by keeping households connected rather than separate as for SHSs, a micro-grid set-up would enable more energy and power provision for all households, particularly if generation was increased. The evolution of a solar nano-grid model into a micro-grid is also possible, and some centralized productive use activities

could be retained at the hub, to reduce the energy and power specification of the distribution lines and equipment to the households.

7.4 Summary

This chapter addressed the question: How can electricity demand data and household demographic data be used to estimate load profiles?

In the first part of this chapter two methods to link electricity data to demographic data were outlined. The first built on the results of Chapter 6, to find the distribution of communal energy modes and average load profiles for different types of households. The second method used the raw electricity data to derive average load profiles and investigate the variance in electricity use between different types of households. Both methods extracted the same trends in electricity use across households for the case study, but the former ensures all data are utilized, whereas the latter is less robust, over-and-under sampling available data. Having linked electricity use to demographic data, the case study results were used to calculate the expected load profile and communal energy modes for a feeder of a micro-grid for a fictional community.

The second part of the chapter showed how the different data types (measured electricity data and aspirational appliance data) gathered in this research can be used together to estimate future household load profiles. Insight about active occupancy in the evening from measured electricity data was combined with data about aspirational appliances and aspirational use patterns. This enables the estimation of load profiles and daily electricity consumption at higher tiers of the MTF, exploring the load growth that could occur given the aspirations of the end users.

8

Conclusions and Further Work

8.1 Conclusions

This research sought to answer the question:

What techniques can be developed to address the challenges of context specificity and replication at scale in the design of electricity access systems suitable for rural, low-income communities in the Global South?

There is a tension in this question between the requirement that electricity access systems are suitable for the specific context in which they are placed, and the need to replicate systems at scale to wider communities. Therefore, the primary research question is addressed through two sub-questions:

1. *To build electricity access systems that are suitable for specific contexts, how can we take account of end-user needs, aspirations, and context in the design process?*
2. *To support replication of electricity access systems at scale, how can we use time series electricity demand data coupled with demographic data to estimate electricity demand for these communities?*

In the process of answering these two questions, five techniques have been presented in this thesis: one technique contributes towards answering the first sub-question, and four contribute towards answering the second. In addition, the results from the case study contribute to the second sub-question.

To address sub-question 1, technique 1 is the Service Value Method, presented in Chapter 3. It outlines steps to gather, interpret, and integrate data on end-user needs, aspirations, and local context, into the design process in a structured and rigorous way. The data-gathering step is designed to gather quantitative data about needs and aspirations of the community, which is then enriched by qualitative data that delves deeper into the practices, norms, and contextual issues around individual and group preferences. This step is carried out in focus groups of community members, in ways that reduce the barriers to the effective participation of end users that may be present in Global South, low-income contexts (see Section 2.1, Chapter 2). For example, pictorial cards are used to represent services that are important to participants to overcome language or literacy barriers, and participants do not require an understanding of technology, or of the design process, to participate.

It is not sufficient just to gather the data required for a design project; a method to integrate it into the decision-making processes in the design is required to ensure the data inform decisions without the influence of designers' biases or assumptions. The Service Value Method provides a rigorous means to integrate the data into the design process, using the quantitative and qualitative data gathered, and including the perspective of the key aims of the project. Key design criteria are selected that reflect the aims of the project, and the services are evaluated against them (structured processes to evaluate services against two example criteria, ease of delivery, and near-term economic benefit, were outlined in Chapter 3). Service maps, showing the popularity data, and how each service stands relative to the key design criteria of the project, enable designers to make decisions about what services the electricity system should support. In this way, the Service Value Method advances the use and inclusion of social science research techniques with full integration into engineering design practice. It is envisaged that it would be used by system developers, whether NGO or private company, as an early step in the process of development of an electricity system. Alongside gathering the important contextual data required, it is also a means of community engagement and fostering of ownership.

Furthermore, the Service Value Method is structured to be easily repeatable across communities, and is a rapid way to include contextual data in preliminary design decisions for electricity access systems. Therefore, it actually addresses the tension in the over-arching research question directly. It is both a means to enable design of suitable electricity access

systems that align with a given context, but is itself easy to replicate and carry out in new communities. The method directly supports replication of electricity access systems at scale.

To further address the issue of replication of electricity access systems at scale, Chapters 4, 5, 6, and 7 address sub-question 2, focusing on estimation of electricity demand for new communities using existing electricity and demographic data. Demand estimation is an essential step in the design process of electricity access systems, and accurate demand estimates are required to cost-effectively size new systems. The aim is to link electricity use to easy-to-measure demographic data about the customer. Then, system developers can estimate demand through a simple survey of the target customers, providing an alternative method to the energy-use surveys that are currently used.

In order to estimate demand for new communities using electricity data coupled with demographic data, it is necessary to collect electricity demand data from existing electricity systems. Time series electricity data are required that accurately depict the electricity demand of rural, low-income communities. A challenge affecting the accuracy of time series electricity data is drift of the real-time clocks in microcontrollers commonly used to monitor electricity use in electricity systems. Without real-time time synchronization during the monitoring period, distributed sensors can experience clock drift, causing time stamps recorded against the electricity data to be inaccurate. In addition, it may not be possible to gather readings of sensor time, and associated actual time, for all the sensors in a system. To solve this problem, Chapter 4 sets out technique 2: a method to estimate rates of clock drift (drift rates) for sensors for which there are no sensor and actual time readings. Through a series of statistical tests performed on sensor and actual time readings for a sample of the sensors, the behaviour of the drift rates is investigated and estimates made for sensors without the time calibration data. This enables post-monitoring correction which increases the accuracy of the electricity data set gathered, so it is fit for analysis.

In Chapter 4, the time correction methodology is demonstrated on electricity data gathered from users of battery-charging systems in two solar nano-grids in rural, low-income communities in Kenya. Battery-charging systems are centralized charging points for batteries, which are the means of electricity distribution to nearby dwellings, and can provide basic electricity access covering tiers 1 to 2 of the Multi-tier Framework for electricity access (see Figure 1.1, Chapter 1). The solar nano-grids were designed and installed as part of this

research, and are the case study for this work (see Section 1.3 in Chapter 1). The battery-charging stations in the solar nano-grids provided electricity to end users' homes via portable household batteries carried between the hub and the household. Use of the household batteries was monitored over five months and, after processing in Chapter 4, resulted in time series electricity data for 51 households, with lengths of data ranging from less than a week to five months. Given the lack of electricity demand data discussed in Section 1.1 of Chapter 1, this electricity data set is an important contribution to the field and enabled the demand estimation analysis in the rest of this research.

The investigation of how to use electricity data and demographic data to estimate demand for new communities focused on two different demand metrics, each applicable to a different electricity system. For battery-charging systems, to size the portable battery and generation capacity at the charging hub, the important demand metric is total daily electricity consumption at the household level. Wired electricity access systems, such as micro-grids, are typically a step up to higher electricity access tiers, and are appropriate for communities that are ready for higher energy and power services. The demand metric required to size the generation, distribution, and potential storage components of wired electricity access systems is the load profile throughout the day. The appropriate system for a particular community will depend on multiple factors including end-user needs and aspirations (as explored through the Service Value Method), and ability- and willingness-to-pay. As both are viable electricity access solutions, to enable rapid replication of either, both forms of demand were investigated.

For each demand metric, the approach taken follows the same generic steps. Firstly, local electricity data from an existing electricity access system are analyzed to understand how electricity is used in the community. Secondly, a link between electricity use and demographic data is found, and this is used to estimate the demand for new communities from demographic data. In order for the demand estimation methodologies to be viable to use in system development, it is important that easy-to-measure demographic data are used in the analysis to provide a link to the electricity data. Although electricity use is complex behaviour to explain, because it is the result of a variety of interacting socio-cultural and contextual factors, the aim of this research is not to use a large number of explanatory variables to explain electricity use. There is a time and financial cost to gathering demographic data. In addition, end users may be unwilling to divulge sensitive information, and even in studies when large numbers of explanatory variables have been used in regression analysis for electricity use,

the majority of the variation in the electricity use remains unexplained (see Section 2.2.2.2 of Chapter 2). Instead, a rapid methodology relying on data measured through a quickly administered survey, such as number, ages, genders, and occupation of household occupants, is of more practical use to system developers.

In the review of demand estimation methods in the literature (Section 2.2 in Chapter 2), a gap was identified in existing methods that use electricity data and demographic data to estimate either single aggregate measures of electricity demand or load profiles. The variation in electricity demand of each customer is not accounted for, as only one aggregate measure or one load profile is predicted for new customers or systems. To address this gap, this research took the approach to retain the variation in electricity use for each household, and examine how demand can be estimated to include this variation. Therefore, to examine total daily electricity consumption, a discrete probability density function based on empirical data is used to describe the electricity use of single or groups of households. Similarly, for load profiles, each household is represented by multiple typical load profiles, rather than just one.

To enable demand estimation for battery-charging systems, technique 3 is the analysis of total daily electricity consumption and the link to demographic data outlined in Chapter 5. Using an easy-to-measure demographic attribute, households are grouped according to type, and discrete probability density functions show the distribution of total daily electricity consumption for each type of household. For a new community, having gathered the appropriate demographic data by survey, the probability density function of daily electricity use can be derived according to the type of households in the community. This can be used to determine an appropriate end-use electricity requirement for households, which is used to size the household battery, and to indicate the electricity requirement at the charging hub.

Chapters 6 and 7 cover a similar process applied to estimation of load profiles for new communities. The first stage of the approach, analysis of existing electricity data, extracts a concise set of typical use patterns from the data. Technique 4 is a process to do this and is presented in Chapter 6. It sets out steps for a clustering analysis to apply a range of algorithms with different clustering properties, and the criteria and method to judge the best solution, which are appropriate for Global South electricity data. The process incorporates a two-stage clustering approach [102] to retain the variation in the use patterns of each household, which has also not been applied to Global South data before. For the first stage, each household is analyzed separately, and clustering is used to identify a number of representative load profiles

that adequately capture the use patterns within each household. For the second stage, all the household representative load profiles are clustered together to give a number of communal energy modes. These are load profiles that describe the range of patterns across the community. The result, for each household, is a distribution of communal energy modes showing what modes the household used, and the proportion of time each mode was used for.

The analysis in Chapter 6 enabled the analysis in Chapter 7, which comprises technique 5 of this research and the final method associated with answering sub-question 2. Households are partitioned into groups according to a demographic attribute, and the distribution of communal energy modes is derived for each type of household. As for the total daily electricity consumption case, demographic data for a new group of households can be used to derive the distribution of communal energy modes for the group. This method enables the estimation of load profiles for micro-grids for new communities, aiding the replication of systems to un-electrified communities.

Alongside the aforementioned contribution of a data set of household electricity data, the communal energy modes derived for the case study are an important result that indicate how electricity for lighting is used in the case study context. In Chapter 6, four communal energy modes were extracted. They show different patterns of electricity use ranging in magnitude from low to high use, and differing in time and duration of use, with one mode in particular representing use into the early hours of the morning. It is unexpected that four distinct modes of use describe the electricity behaviour given the electricity data set is from just one service that is a basic need (lighting) and that it is provided on a flat tariff model. In addition, each household used the same specification of lighting service, and the dwellings were similar in size and layout. This result suggests that different norms and practices of household occupants cause different ways of using electricity, even for households in similar circumstances, and for a simple service such as lighting.

A second key case study result relates to the demonstration of the demand estimation analysis using the demographic attribute ‘number of occupants’ of each household. Both the analysis of total daily electricity consumption and load profiles showed a trend that larger households use more electricity than smaller households, and more frequently use electricity later into the evening. A two-fold difference in peak average power for lighting was observed between the smallest household category (1-4 people) and the largest (9-12 people). This result indicates that the number of occupants in a household impacts the

norms and practices that influence electricity use, causing a notable difference in lighting electricity use. Looking to the future, when more electricity services are accessible, this gives an indication that similar differences in practices and norms would cause a variation in use of higher energy and power services within a community.

The potential for demand growth over time is an important consideration for electricity access systems, and the second half of Chapter 7 explores what future load profiles could look like and how daily electricity consumption is affected as end users access more appliances. Rather than estimating electricity demand at one point in time, this builds on the demand estimation techniques to indicate how demand could increase as the economic circumstances of end users changes and they are able to afford more electricity services. Electricity use data and data about aspirational services as gathered by the Service Value Method are combined to estimate a future load profile for an indicative household. This is a combination of the tools that address context specificity and replication of electricity access systems at scale, showing how the techniques can work together. While demand estimation can be based on existing data, suitable aspirational services can be included in the load profile to indicate future demand for a specific context.

The techniques presented in the thesis could be applied by a company/NGO/utility to contribute to the design and deployment of a financially viable electricity access system that provides economic benefit. By ensuring the system meets end-user needs and aspirations by incorporating the Service Value Method into the design process, the system will deliver useful services that improve quality of life and enable income-generation. By more accurately estimating electricity demand using the demand estimation techniques and household energy data set, uncertainty in demand is reduced. This means the system can be sized so that electricity supply more accurately meets demand, reducing system cost compared to over-sizing to cope with unknown demand. The cost-saving can be passed along to end users through an affordable payment model, and reduced capital expenditure can contribute to the system recovering costs and being economically viable for service providers.

8.2 Further Work

This section presents four possible research questions that build on this research.

How does the demand estimation methodology presented in this research compare in accuracy to energy-use surveys and actual measured electricity use?

This question is the next step required to evaluate the demand estimation methodologies developed in this thesis. Comparison should be made with actual measured electricity data, and also with energy-use surveys, as this is a common method currently used by system developers.

To minimize costs, ways to investigate this using existing electricity data should be explored first. This requires a training set of data and a test set, each comprising electricity data and associated demographic data for a group of households. For example, if there is another project with electricity data and household demographic data available, the electricity demand could be estimated using the case study used in this thesis, and compared to the measured demand. The comparison against energy-use surveys depends on the availability of energy-use survey data. If, as in Blodgett et al. [44], there is existing energy-use survey data from before the electricity system was constructed, the demand estimate derived from this should be used. However, if energy-use surveys were not used, or the data are unavailable, new surveys can be conducted to gather data on the current electricity use of the ‘test’ community as in Hartvigsson et al. [86]. The former circumstance is a more realistic test of the two demand estimation methods, as it reflects what occurs in the field in reality. However, the latter circumstance still allows for an interesting comparison between the method presented in this research and self-reported electricity use.

A full field-trial could also be used to investigate this question. Energy-use surveys and the demand estimation methods in this research could both be used to estimate demand for a new community. Once a system was built, both estimations could be compared to actual measured demand to see which was the more accurate. The challenge of this approach is the financial cost of new system development and deployment.

How does the electricity demand of rural, low-income users compare across geographical locations, cultures, and electricity systems, and what demographic variables are important to explain electricity use?

The methods developed in this research can be used to analyze data from different electricity systems, deployed in a range of geographical locations and different cultures. A comparison between different areas, cultures, and systems, is important for understanding how electricity demand varies across different contexts. In addition, it would enable a library to be built up, cataloguing the results of the various case studies. The case studies

would be tagged using location, type of electricity system, services enabled, and a short description describing the culture and context of the study (we recommend use of the Energy Cultures Framework for this). It would not be possible to identify the exact households from these data, and it would adhere to data privacy and protection legalities and concerns. The associated electricity data would be estimated discrete probability density functions constructed using the methods presented in Chapter 5, associated with different household types. It would also consist of distributions of communal energy modes, also associated with different household types, and processed using the methods presented in Chapters 6 and 7. This library would then be available to stakeholders such as system developers, government, NGOs, and researchers. It would be useful for electricity planning at a range of scales, from assessing overall electrification strategy on a macro-level, to determining the type of electricity access system that is suitable for a specific location, to sizing the system according to expected demand.

What is the least-equipment lighting service for the solar nano-grid that still meets end-user lighting needs?

Examining the lighting data at a finer time resolution would allow the interpretation of the data in terms of the number of lights required, rather than the magnitude of energy/average power (the approach in this research). This would enable optimization of the lighting service in terms of specifying the minimum number of light-bulbs that meet the needs of the end users. It would also be helpful to conduct activity surveys to understand the task-lighting energy requirement in the households, to assess if this could be included in the solar nano-grid lighting service.

How do energy norms and practices differ across households, and how does this effect the demand for electricity and what are the implications for future electricity demand?

Understanding how practices differ across households, and why they differ (the underlying norms), will give insight into the behaviour behind particular electricity use patterns. This is useful to ensure electricity systems work within end-user norms and practices, and also enables better prediction of how electricity demand might vary at higher tier and service provision. For example, if households with children, that also tend to be larger households, tend to be occupied during the day for child-care reasons, this will indicate that daytime use can be expected of higher tier electrical services. Provision for this can then be made in the system design.

It is proposed that this is first tackled with the case study communities in this research, given the existing electricity data and the progress in analyzing it. Specifically, understanding the practices that are behind the communal energy modes would allow for better design of the lighting electricity system. For example, the system could be tailored to provide task lighting or background lighting, depending on the function that is important to the end users.

Furthermore, it would be interesting to understand why some households use late-peak mode, which showed late lighting electricity use into the evening and early morning. The research in this study showed that it does not appear to be correlated with the number of occupants in the household. However, it represents an important edge case of demand where electricity is required later than usual, and occupants remain active later into the night, which has ramifications for current electricity demand and the prediction of future electricity demand. Finding out why late-peak mode occurs in these case study communities will indicate what factors might indicate its prevalence in other communities. If, for example, it is connected to the occupants' location of work, late-peak mode might be a more common in communities where many people work outside of the village and return later at night, which could be picked up through surveys. This would then inform the designer that time and duration of use may be later in the evening and longer than expected.

This research could proceed through targeted surveys of households who use late-peak mode, asking about the time that activities in the household usually take place. In addition, data should be gathered on place of work and time and means taken to travel. As discussed in Chapter 3, there are challenges in ensuring surveys are administered correctly, though the barriers are not insurmountable given sufficient time investment.

Appendices



Supporting Material for Service Value Method

A.1 SVT Equipment

The following equipment is suitable for one SVT with 10 participants.

- Service chart: A1 laminated card with empty 12 box grid, placed centrally for the group to see
- Service cards: Laminated cards showing pictorial or symbolic representations (that are contextually appropriate) of different household and community services, and extra blank cards for drawing services not anticipated
- Non-permanent marker for drawing/writing extra services
- Blue-tac for fixing service cards to the service chart
- Service chart index cards: 10 A5 versions of the service chart for participants to make individual selections
- 10 pencils for participants
- 200 counters (10 different colors, sorted into groups of 10/20 depending on the preference indication method used)
- Recording equipment (audio/visual/both)

A.2 SVT Implementation Recommendations

People:

- Multiple focus groups should be run for the same community, separated according to demographic traits such as gender and age, to a) get a cross-section of community viewpoints and b) enable those with less social power to speak unhindered.
- For each focus group, 10 participants and two facilitators is a good balance between hearing a range of opinions and making group discussion possible and inclusive.
- Local power and governance structures should be respected and managed sensitively in the choosing of focus group members, given that the presence of strong influencers that might inhibit participation of other participants, but that community engagement may only be possible with their support.
- If a translator is required, they should be someone external to the community who is neutral to the project, chosen with cultural norms and social power dynamics in mind.
- Facilitators should be chosen in consideration of cultural norms that might limit participants' freedom, willingness, or comfort to speak. For example, in a patriarchal society, female facilitators should facilitate a focus group of female participants.

Process:

- Obtain informed consent from all participants for their participation in the focus groups and for using the resulting data.
- Respect cultural and social norms while designing focus groups and selecting participants and facilitators.
- Participants' expectations should be carefully managed regarding the aims of the focus group to avoid possible misunderstandings (e.g. that the exercise is not a promise of requested services).
- In organizing the focus groups the aims of the focus group should be articulated to the participants to emphasize that their general opinions and practices are of interest, rather than just those about energy.
- The translator should be briefed on the exercises before the focus group.
- The service chart should be placed on a central table and participants seated so they can see the chart and can be fully included in the exercises.

- Session rules should be agreed at the beginning of each session to enable smooth running and maximize data collection e.g.: only one person speaks at a time; all opinions are valid.
- Facilitators should be aware of their own biases and expectations and minimize the influence these have on their interpretation of responses.
- The goal of the facilitators is to create a safe, open space for participants.
- If there are mixed language skills in the focus group, it can be challenging for everyone to follow the discussion. Decide if the translator will back-translate any comments in English to the participants' language so that everyone understands.
- Photograph the service chart at the end of steps 1 and 2 and collect service chart papers.
- If possible record session audio and get transcriptions, particularly for step 3 of the SVT.

A.3 Summary of SVT Discussion for Kenya

Tables A.1 and A.2 summarize the key points from the discussions in the SVTs run in the Kenyan villages, for household and community services respectively. Shaded entries indicate that the service was not suggested by any groups in that community during the SVTs, and blank entries mean there was no particular discussion about that service.

A.4 Assumptions about Suggested Services

Table A.3 details what each service was assumed to consist of in terms of electrical products, infrastructure and additional complexity. The electrical power of each service are noted and come from existing products or from a catalog of productive use appliances compatible with power generation by PV panels [149].

Table A.1: Summary of SVT Discussion for Household Services, Kenya

Current situation			Why is the service important?	
Service	Lomolo B	Echariria	Lemolo	Echariria
Torch	Common	Common	Going to latrines at night. Walking around the village at night.	In case run out of kerosene. Going to latrines at night. Security when walking at night
Fan	Not common		It is very hot in the day.	
Radio	Common	Common	News (local and global), educational programs to advance themselves, particularly about agriculture, political programs, music	News (local and global), educational programs, particularly about agriculture, music.
Electric Cooker	None	None	Have to collect firewood, there is a shortage of it and it takes a long time	Firewood is hard to get - buy or collect it. Collecting it takes a long time, usually do this once a day. Would save time that could be spent working. Charcoal is expensive. Both have ill health effects and are dirty.
TV	Only 1 in the village	People in lower villages have TVs	News (local and global), educational programs, entertainment, sport.	(E youth: colour TV) News (local and global), educational programs (agriculture and general), entertainment (soaps, comedy)
CD/DVD Player	None	None		Entertainment Listening to sermons
Solar Powered Shower		None		Saves time. Warm showers are better
Kettle	Heat water on fire	Heat water on fire	Heat water for washing. Boil water for making tea.	Boil water before drinking. Warm water for washing.
Iron	Charcoal iron	Charcoal iron	Charcoal is difficult to get and is smoky.	Charcoal irons are heavy, dirty, not safe, time consuming. Looking neat and smart is important. Saves time compared to charcoal iron.
Lighting	Kerosene lamps or wood fires.	Kerosene lamps	Children can't read and people can't work into the night. Children studying. Able to see to cook and prepare food. Reduce accidents (cooking, snakes). Kerosene is bad for health. Wash clothes and hold meetings at night.	Children studying. Able to see when cooking in evenings so there are fewer accidents Have to buy paraffin regularly.
Computer		Use a cyber cafe in Nakuru.		Research, studying and information finding (medicine, agriculture). Advertise their business, looking for jobs. Typing letters. Watching programs.
Washing Machine	None		To have a laundry business.	
Mobile Charging	Charge during the day.	Charge in the village or at a neighbors house	Want to be able to charge at night Phones are important for: communication with friends and relatives, communication when away from the village.	
Fridge	None	None	To store milk To preserve food.	Vet wanted a fridge for vaccinations. Food storage - saves time shopping, as one has to go less frequently.
Welder	Welding services in the local town		To save time travelling to the local town. To fix farm tools as they are all farmers. To have businesses: fixing bicycles and motorbikes; making iron windows.	
Shaver				To keep children clean

Continued on next page

<i>Continuation of Table</i>				
	Current situation		Why is the service important?	
Service	Lomolo B	Echariria	Lomolo	Echariria
Hairdryer	None	None	For beauty and for starting a salon business.	For looking smart for interviews. To manage their hair. Family members would like it.
Sewing machine		None		For starting a tailoring business.
Incubator	None	None	Poultry farming is a business opportunity.	(E women: includes lighting and heating equipment for chicks post-hatching) Can earn money from poultry farming. Can use the money for school fees. Protein is good for the health.

Table A.2: Summary of SVT Discussion for Community Services, Kenya

Current situation		Why is the service important?		
Service	Lomolo B	Echariria	Lemolo	Echariria
Health center	No health center in the village, the nearest is in a town 30 minutes away.	There is a dispensary. It has no power, water, fridge, or laboratory. Have to go to local big town for health services. There is a private clinic in a lower village but still cheaper to go to the big local town. Can get painkillers from nearby villages.	Biggest health needs are: maternity, malaria, typhoid, TB, snake bites, gastro-diseases.	Currently have to travel far to access service. There are not enough medicines. Children have to be taken out of the village for vaccinations.
Security Lighting	None	None	(Elders: included police) Moving about safely in the night. Police required to address conflicts about land ownership and access that sometimes occur with people who were already in this area.	(Youth: included police) Theft is a problem. Important for walking around safely at night. Businesses need security to thrive. There is an alcohol problem that makes places unsafe.
Pre-primary Education			Children start learning earlier and parents have more time to work.	
Water	Collect water from a dam, but it is bad quality (also used by livestock). Collected daily by mother and/or children, 1-3 hours. It is not easy to sink wells here as the water table is very far down.	Pump for borehole is broken. Some people get water from the river, collecting daily for ~1 hour, but there is risk of disease (cholera, typhoid) and sometimes it dries up. Some people buy water from others who collect rainwater or have piped water supply.	Clean water is needed for drinking, washing and for livestock. Currently collecting water takes a lot of time.	See left.
Education	There is a primary school but no electricity or facilities. For secondary school children walk far.	Primary and secondary schools, grid electricity but only lighting.	Primary school should have electricity, food, sport facilities. Nursery school would free up time for working. Secondary school required (with laboratories for science, library; dining hall could be used for community meetings). Technical college required for computer training, plumbing, carpentry, masonry, agriculture, tailoring, nursing, teaching.	Children can learn trades to make money. Problem with children dropping out of school, they need to be kept busy. Youth themselves would like more education. Polytechnic, in addition to left list: (driving, welding, mechanics, hairdressing, catering)
Market Place (permanent building with lighting)	Don't have a permanent market building	Don't have a permanent market building	Building and selling produce	Buying and selling produce
Carpentry				
Improved Roads	The community is in the middle of a sisal plantation with bad quality dirt roads to the local town (30 minutes).	Roads to the village are bad: 'Shambles'. Bad in wet weather.	So that people can come in and out easily. To access the health center at the local town. To transport farm produce. To access work opportunities in the local larger town (1.5 hours away). To allow teachers to come in to work at the school. To save time in all the travel they do.	(Women: included an improved bridge, and transportation methods - so represented general access) Required to transport goods. So people can come and go easily. Be accessible for potential buyers and sellers. Travel to big town for business.

Continued on next page

<i>Continuation of Table</i>				
Current situation			Why is the service important?	
Service	Lomolo B	Echariria	Lemolo	Echariria
Welding				Can provide jobs
Crop storage (elevated dry storage space)				
Salon and Barber	None	There are salons in the village but none have electricity		See left.
Community center				Could run business from a community center that had access to computers, TVs.
Irrigation	Rainfall is erratic and not reliable. Droughts are frequent.	Soil is very dry. Water is scarce.	Want to sell excess produce at local town.	Could produce extra food to sell
Grinding Machine (to make flour from maize)	A couple of diesel machines in the village.	Current machines use diesel, or are manual and hard work.		Can provide jobs.
Shave Cutter (makes animal feed)				
Sports Playground	None.		For community sports, and for people to develop sporting talent.	
Bank	None.		Access to a bank account. Access to loans and micro-finance.	
Church	Currently people meet outside underneath trees for church.	None of the churches have electricity.	See left.	Want amplification of instruments and voices.
Tractor and pesticides				
Sanitation		Need safe and hygienic pit latrines, but the ground is rocky, it is difficult to dig latrines.		
Ambulance	None. In emergencies people go on motorbikes to the local town.	None.		
Improved mobile reception	Mobile phone reception is currently patchy.		See left.	
Financial support for the elderly	Currently the elderly cannot access financial support from the state. In the place they lived before being displaced, they were able to access this support.		See left.	
Alternative mobile phone operator				

Table A.3: Assumptions made about Services

Service	Electrical products included/ notes	Power estimate	Reference	Infrastructure component?	Additional Complexity?
Torch	Rechargeable	<10W	INFRAY Rechargeable 800 lumens LED Torch		
Radio		2W	Page 190		
Lighting	Assume 3 3W LEDs for household	9W	GRV E27 6-5730 SMD LED Globe Bulb		
Fan		5-20W	Pages 172-174		
Mobile phone charging (household)		5-10W	Nokia AC-60X Universal Fast Micro USB Charger		
TV		8W	Pages 191		
DVD player		10W	PANASONIC DVD-S500 progressive scan DVD player		
Computer	Assume mobile network used for Internet connection	30W	Page 168		
Printer		20W	Pages 177-179		
Shaver		10W	Anself Hair Clipper Model: CJ-907		
Security light (household)	Two lights outside house	60 W	MEIKEE 30W LED Motion Sensor Flood Light		
Fridge	For vaccines:	50W	Dulas solar direct drive vaccine refrigerator VC50SDD		
	For food storage:	45-135W	Pages 102,109,114		
Sewing machine		30-100W	Pages 144, 146, 147, 150		
Incubator	Depending on size	80W-175W	Pages 36, 38		
Electric cooking/ rice cooker	Assume equivalent to a multi-cooker	900W	Russell Hobbs Multi-Cooker 21850, 5 L		
Rice husking		250-375W	Pages 82-84		
Grinding machine		750W	Page 76		
Shave cutter		2.2kW	Whirlston Mini Chaff Cutter 9ZP-0.4		
Power tools	Assume one tool used at a time.	400W	Page 161		
Kettle		400W- 2.2kW	Page 132 LOGIK L17JBS17 Jug Kettle		
Iron		150W- 1.1kW	Pages 221,222 ESSENTIALS C12IR13 Steam Iron		
Hairdryer		400W- 2kW	Page 227 REVLON Harmony 2000 Dry Style Hair Dryer		
Welding		3.5kW	Trueshopping AC Arc Welder 100 Amp		
Water	Dependent on: location, ground type, water table. (Place in highest power category.)	0.9-1.4kW	Pages 60-62	Possible bore hole, possible pumping over distance, possible filtration/cleaning.	

Continued on next page

<i>Continuation of Table X</i>					
Service	Electrical products included/ notes	Power estimate	Reference	Infrastructure component?	Additional Complexity?
Solar powered shower	Depends on access to water.	As water		As water + additional distribution	
Washing machine	Depends on access to water. Additional for machine.	As water 70W	Page 219	As water + additional distribution	
Household water supply	Depends on access to water.	As water		As water + additional distribution	
Irrigation	Depends on access to water.	As water		As water + additional distribution	
Pre-primary education	Lighting: assume 10 inside lights (30W) and 2 security lights (60W)	90W	See Lighting and Security lighting (household)	Building	Teacher, association with national education service
Market place	Lighting: assume 10 inside lights (30W) and 2 security lights (60W)	90W	See Lighting and Security lighting (household)	Building	
Mobile phone charging center	Assume ~20 phone chargers	150W		Assume occurs in the solar hub	
Security lights (community)	Assume 6 30W lights around central hub/central street	180W	See Security lighting (household)		
Sports playground	Assume 6 30W lights around pitch	180W	See Security lighting (household)	Sports field	
Health center	Assume vaccine fridge, computers, 20 inside lights, basic medical equipment (200W+), 2 security lights	500W +	See Fridge, Computers, Lighting, Security lighting (household)	Lomolo, Faitang, Baroihati: Building	Trained medical professional, association with national health service
Improved roads (street lights)	Assume 10 30W lights (depends on space to be lit)	300W	See Security lighting (household)		
Education (incl. Religious school)	Computer server for 6 computers (300W), 6 computer monitors (70W), 20 inside lights, 2 security lights	490W	Computer server: Page 169 Computer monitor: Page 168 Lighting: see Lighting and Security lighting (household)	Lomolo, Faitang: Building for secondary school Lomolo and Echariria: Building for polytechnic college Baroihati: Building for primary school	Teachers, association with national education service
Carpentry	Power tools (assume one at a time), 10 inside lights	430W	See Power tools	Building	
Salon and barber	3 shavers, 1 hairdryer (400W), straightener (50W), 10 inside lights	510W	Straightener: GLAMORISER Touch Control GLA023 Hair Straightener See Shaver, Hairdryer, Lighting.	Building	
Community center	Computer server for 6 computers (300W), 6 computer monitors, 1 TV, 10 inside lights, 2 security lights	490W	Computer server: Page 169 Computer monitor: Page 168 Lighting: see Lighting and Security lighting (household)	Building	
Garment factory	Assume 2 industrial sewing machines (230W), range of smaller machines (30-100W), 20 inside lights	800-1.5kW	Industrial sewing machine: page 148 See Sewing machine, Lighting	Building	
Church	Assume speaker (1kW), Mixer (55W), 3 microphones (10W), 10 inside lights	1.1kW	Electro-Voice ZLX-12P 12 inch Powered Loudspeaker Yamaha MGP16X 16 Channel Premium Mixing Console Shure BLX24R/SM58 Handheld Wireless System See Lighting	Lomolo: Building	

B

Additional Electricity Data Processing Steps

Contents

B.1	Unrecorded Time Calibrations within Monitoring Period	210
B.2	Battery RTCs were Reset to 1970	212
B.2.1	Multiple Resets	212
B.2.2	Calibration to 2016	213
B.3	Microcontrollers Logged Data with Inconsistent Frequencies	214
B.3.1	Gap-filling Algorithm	215
B.4	Addressing Logging Disruptions During Charging	217
B.4.1	Identification of Logging Disruptions	218
B.4.2	Typical Charging Profiles	218
B.4.3	Filling in Missing Charging Data	221
B.5	Power and Duration of Use Limits	221
B.6	Testing Loading Conditions to Ensure ‘Normal’ Use Data	224
B.6.1	Monthly Electricity Use Variation	225
B.6.2	Day of the Week Electricity Use Variation	227

These are the additional processing steps referenced from Chapter 4, Table 4.2, which were necessary to process the electricity data collected in the case studies to useful household specific time series electricity data.

B.1 Unrecorded Time Calibrations within Monitoring Period

In addition to time drift, the time of some batteries was calibrated to the actual time while they were in the field. When exactly this occurred was not recorded, but was possible to detect

through examining the state of the battery. The calibration could then be taken into account in the time correction process. In two cases the time was calibrated incorrectly, without adherence to the 24-hour clock format. These two cases needed care to interpret what had happened and set time points to calibrate the data back to the actual time.

Time calibrations within the monitoring period usually occurred when visitors to the site, either from the academic team or from the battery manufacturer, updated the time as part of a repair or software update. The battery time readings were initialized to GMT time, so appeared approximately 3 hours off the actual time in Kenya to observers, who then changed the time to local time. Time calibrations are frequent between the 5th and 7th September 2016 because battery manufacturer personnel visited both sites for software updates and technical fixes. Less frequently, the technicians updated the time after repairing a battery. These calibrations are partly why the sample of batteries with measured drift rates between time points (2) and (3), $d_{2,3}$, is small in size. It is not possible to calculate the rate of drift over a period where the time stream has been interrupted in some way.

The first step was detection of when a time calibration had occurred. An automatic check of the battery state was used to search for state ‘7’, which occurs when the time of the battery is updated. The state 7 check found all the battery time readings that registered a state 7, and manual observation was required to pick out the exact calibration point. The exact calibration point was possible to distinguish from a set of State 7 readings because a battery could only be calibrated to the nearest minute. In terms of the time correction procedure, the calibration point become both reference time readings T_a and t_a in equation (4.6) for the section of data that proceeded it.

Where the time was incorrectly set after a repair or modification, neglecting the 24-hour format, the time calibration point detected by the state 7 check was modified. 12 hours were added to correct for the fact that during the failed time calibration, the hours of 13-23 were referred to as 1-11. Otherwise time correction proceeded as described above.

Time calibrations were accurate to the nearest minute, as this was the input time resolution available for the batteries. In terms of accuracy, the calculations in Section 4.4.4 are applicable and are upper bounds, because in all cases where there has been a time calibration, the time span that follows will be shorter than the span considered in that worst-case calculation.

B.2 Battery RTCs were Reset to 1970

For 49 out of 137 of the batteries, the RTC had been reset to 0 in UNIX time (00:00:00, 1st January 1970) and not calibrated again to the correct time for significant periods of data logging. This reset was caused by disconnection of the battery microcontroller from the PCB while the microcontroller was switched off, which broke the connection to the small battery designed into the PCB specifically to power the RTC during switch off. This likely happened during a repair of the battery if the microcontroller was lifted out to get better access to the PCB. The data sets from batteries with significant data dated to the 1970s are classed as ‘reset’ data sets, and this section outlines how the 1970 section of data for each one was dealt with. Where a battery was reset multiple times, there was usually another use profile recorded over the top of an older one in the same csv files. The data sets needed to be cleaned up so they only represented one use profile, and where possible, calibrated to 2016.

B.2.1 Multiple Resets

A major challenge in terms of calibrating the 1970 data to the correct 2016 time was that in some of the reset data sets the time was set to UNIX 0 multiple times, so there was one or more usage patterns on top of another. When a battery is reset more than once, data points append to existing csv files for that particular hour, or create a new file if there isn’t one in existence. When multiple resets occurred and logged for significant periods of time, the time streams of use data are essentially entangled and it is not trivial to identify which recordings belong to which time stream. Different csv files had a different number of repeated logging streaks, depending on the use of that battery.

To find out which data sets and sections of data sets were possible to use while reasonably confident that there were no entangled time streams, an automatic check was used to count the number of time streams present in each csv file. This showed which csv files had multiple time streams and for how long the multiple time streams extended through the data set. The results were used to classify the reset data sets depending on the severity of how the multiple reset issue manifested and determine what sections of data were acceptable to carry forward.

Untangling the time streams was out of the scope of this project, so where the process detected more than one time stream in a data set, the section of entangled time streams was removed entirely. Removing the entangled section of the data is not without uncertainty that

Table B.1: Classification of reset data sets according to the number of resets that occurred

Category	Number of data sets
No usable data – all entangled	12
Ignored as data are insignificant	16
One reset only	6
<i>Data sets with multiple resets: Duration of entangled data</i>	
Within first hour	7
Between first hour and end of first day	4
Between days 2-10	5

the remaining data are truly free of entangled time streams, as it relies on the assumption that as soon as multiple time streams stop being detected, they no longer exist. This may not be true if the different time streams logged battery use for different hours and so logged in different files, never over-lapping. This approach therefore risks leaving in an unknown portion of entangled time streams. To mitigate this risk, heat maps were made of the shortened data sets and carefully inspected to detect if any patterns of use seemed to be duplicated at different times.

Applying this process to the data, Table B.1 summarizes the categories the reset data sets were sorted into, and the number of data sets that fell into each category. For 12 data sets there were long entangled time streams without a significant portion of data from just one time stream. These data sets were put aside as not usable. For 7 batteries there was no evidence of multiple resets, but on inspection of the heat maps one of them was discarded as having insignificant data. The remaining data sets did have multiple time streams, with durations from just within the first hour to 10 days. In each of these cases the entangled section of data was removed, and upon inspection of the heat maps to check for further entanglement, 15 more data sections were found to have insignificant data and were also discarded. This process left 21 of the original 49 data sets with useful 1970 data to contribute to usage profiles.

B.2.2 Calibration to 2016

Out of the 21 data sets with usable 1970 data, only 5 of those had time point readings of battery time and actual time that could be used to calibrate the 1970 data to 2016. For 8 of the data sets, there was no way of knowing or making a reasonable guess of when the 1970 data slotted into the 2016 data, and thus no way to calibrate them. For the remaining 8 cases, inspection of the heat maps for the 1970 and 2016 sections indicated it might be

Table B.2: Calibration status of reset data sets

Category	Number of data sets
No significant data that are not entangled	12
Ignored as data are insignificant	16
Impossible to calibrate	12
Calibration possible with recorded time points	5
Guessed time point data and got realistic results	4

possible to make an informed guess of a time point to calibrate the 1970 data to 2016. Having estimated calibration time points, carried out the time corrections, and plotted the results as heat maps, four cases were found to look realistic, and four were found not to. The four that looked realistic were kept, and their existence in the final household power profiles is noted in Appendix C. Table B.2 summarizes the calibration status of the 49 data sets.

B.3 Microcontrollers Logged Data with Inconsistent Frequencies

The frequency at which data points were written to the SD card by the battery microcontroller varied in the data in two respects. Firstly, there was a set recording frequency that changed after a couple of weeks of deployment when a software update was implemented, changing data logging from every 8 seconds to every 4 seconds. Secondly, the microcontroller often failed to maintain the set recording frequency regardless of what it was meant to be. Logging slowed down when particular states occurred in the battery, was erratic around switch on and off, and also varied whenever a battery went through a change in state, for example, from charging to not charging. Data sets with a consistent frequency were necessary to enable the plotting of heat maps, and were desirable for later analysis.

To address this, each battery data set was re-indexed on a per second basis, and an algorithm used to fill in the gaps between readings, detailed in the following section. The algorithm was constructed from detailed examination of the raw data, to work out whether the gap between two readings should be filled, or whether the battery was in fact switched off during that time. The algorithm was used on each csv file, to determine whether to forward fill between each line of data, and also when csv files were concatenated together, to determine

the forward filling strategy between csvs files. This processing step comes after the data timestamps have been corrected for the drift of the RTC.

B.3.1 Gap-filling Algorithm

The gap-filling algorithm takes as input two adjacent timestamps, t_i and t_{i+1} , calculates the time between them, g , and evaluates g to determine what has happened and how far to fill the data values forward. Figure B.1 shows the algorithm. The method of forward filling uses a hold on the previous data readings, propagating those values as far as determined by the algorithm. This approach was chosen because the majority of forward fill events concerned very small changes to the data readings, so a linear interpolation was deemed unnecessary. In any case, data readings fluctuated by the order of 1% even when expected to be constant, showing noise associated with the measurements. Thus it was not required to use a more complicated forward fill procedure than a simple hold.

The algorithm compares g to the gap mode, which should be the set recording frequency of the microcontroller. If g is larger, the algorithm categorizes the reason for the larger gap and dictates what action should be taken. The first check is for a known issue of the battery entering hibernation state in error during charging, marked by battery state 12, which caused logging to occur precisely every 2 minutes. The battery is not actually turning off, so the algorithm dictates that the whole gap be filled.

The second check of the battery state is checking whether the battery has been switched off, characterized by a battery state reading of 0 when the battery is switched on again. If the battery has been switched off, the data value is forward filled into the time interval that equal the gap mode.

The third check addresses another known slow-logging occurrence, which occurs in the first minute after the battery is switched on. This is accomplished by again examining the state readings of the battery to establish that, if they are both 0, the battery is in the process of turning on. However, logging was delayed in switch off by approximately 20 seconds, so if the size of the gap is larger than 30 seconds, the battery is assumed to have been switched on and off quickly, which occurred when users checked batteries for their SOC.

The fourth and final check for why a longer than expected gap has occurred is based only on the size of the gap, because there were no other detectable reasons for slow logging. The cut-off is set to 1 minute, and larger gaps are assumed to be a switch off, while smaller gaps

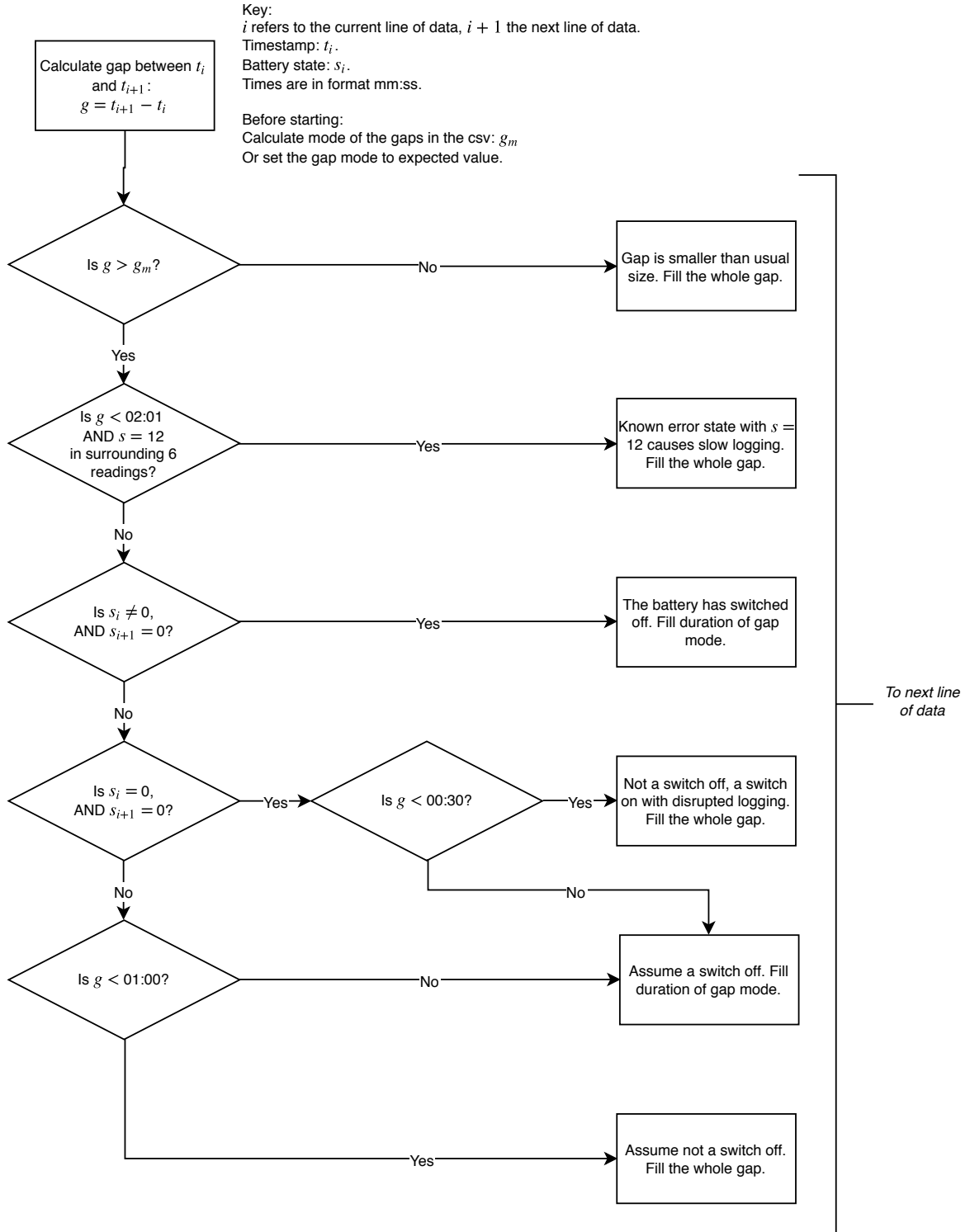


Figure B.1: Gap-filling algorithm.

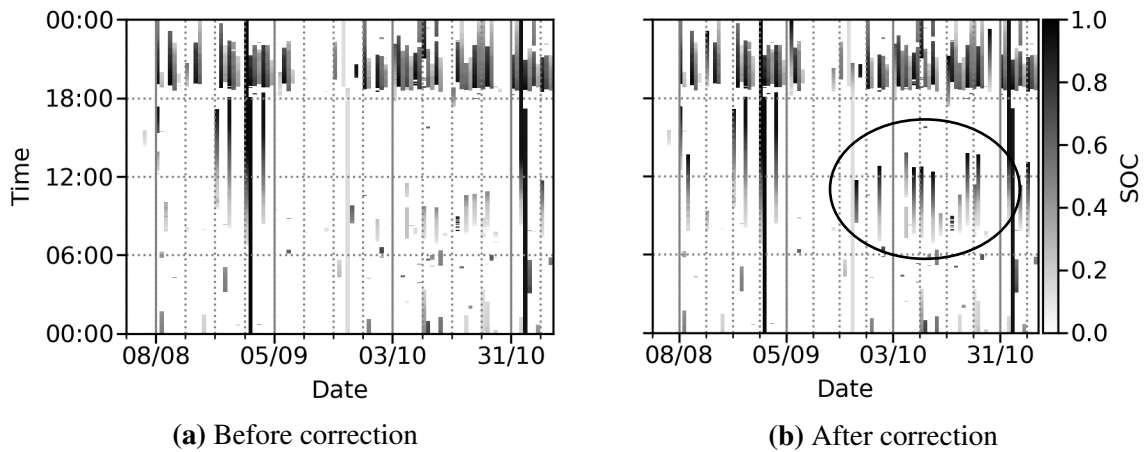


Figure B.2: Heat maps of SOC for one household profile, showing discontinuities in data logging (B.2a) and correction (B.2b).

are assumed to be some other issue that was not detected. In the absence of more information, there was no other check that could be made to infer what was actually happening.

The output of this processing step were re-indexed battery profiles on a one second basis.

B.4 Addressing Logging Disruptions During Charging

Most batteries experienced failures in data logging during charging at some point, meaning that an inaccurate representation of charging behaviour was captured. To get a more accurate representation of charging energy, typical charging profiles were found by examining charging events that were fully logged, and were superimposed onto the disruptions.

Figure B.2 shows heat maps of SOC of a battery before and after logging disruptions have been corrected. The discontinuities in shade on the heat maps show where data stopped being logged during charging of the battery. According to the manufacturer, after a certain software update the microcontroller logging function could crash during charging, switching on again when charging stops. We were also informed that the SOC measurement was unreliable, which became evident during the work, with the SOC seen to jump $\pm 20\%$ during a change of state of the battery (for example, going from charging to not charging), and to show smaller fluctuations at random.

Three steps were required to correct the logging disruptions, as described in the following subsections. Firstly, all the logging disruptions in the data were identified. Secondly, complete logging events were used to give typical charging gradients of SOC and current. Finally,

the typical charging gradients were superimposed onto the discontinuous logging events to give more accurate indications of charging energy and behaviour.

B.4.1 Identification of Logging Disruptions

Discontinuities in the SOC value were used in an automatic check to identify where logging had stopped. Four criteria were used in the algorithm to identify the disruptions, considering the difference in SOC readings i and $i + 1$:

1. Difference between SOC readings was larger than 0.1
2. Time between the readings was greater than 2 minutes
3. SOC reading jumped up rather than down
4. Battery was in a charging state before or after the discontinuity

The first and second constraints put a limit on the size of fluctuation and the time span between the two values. Smaller fluctuations and those that occurred over a shorter time span were ignored because they indicated the fluctuations that seemed to occur randomly in the SOC values, and did not constitute logging stopping for a significant period of time. The third and fourth constraints ensured the discontinuity was associated with the known charging issue, and wasn't due to the jump down that occurred when charging was first instigated. This process gave 985 logging disruptions during charging across all the profiles.

B.4.2 Typical Charging Profiles

A typical charging profile was sought that could be superimposed over the discontinuities. For simplicity, the typical charging profile was assumed to have a SOC profile with a constant gradient with respect to time, corresponding to a current with constant gradient, until the battery was fully charged at SOC = 1. This ignored the tailing off of SOC and current that typically occurs around 80% SOC. However, given the unreliability of the SOC reading, and the differences in when the SOC gradient starts to decline between batteries, this is a reasonable assumption to make.

The methodology to calculate a typical charging profile involved picking out all fully-logged charging events that resulted in an SOC above 0.7. A linear least squares regression model is used to find an equation to describe the SOC and current of each individual charging event over the 'constant current' charging section (where the current is almost constant for

Table B.3: Average gradients for SOC and current for each charging current category

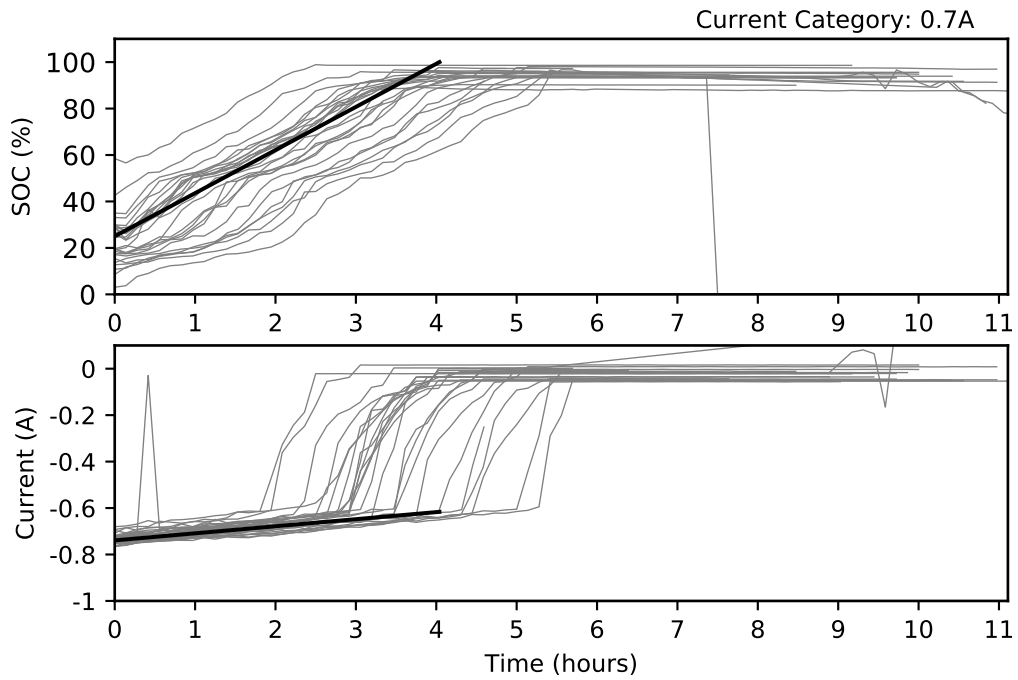
	0.7 A	0.5 A
Number of charging events in sample	260	182
Average SOC gradient (/hr)	0.19	0.12
Average current gradient (/hr)	0.03	0.01

the bulk of the charging, but in fact has a very small positive gradient with respect to time as acknowledged in this model). Equations (B.1) and (B.2) show the equation for SOC and current for charging event i , where m is the gradient and c is the intercept. The SOC and current gradients of each individual charging event are averaged to find the SOC and current gradients for the typical charging profile, as shown by (B.3) for the SOC current. The average current gradient is calculated in the same way. Calculating an intercept c for the typical charging profile is unnecessary because when superimposed into the data, the typical profile will simply start from the last data reading.

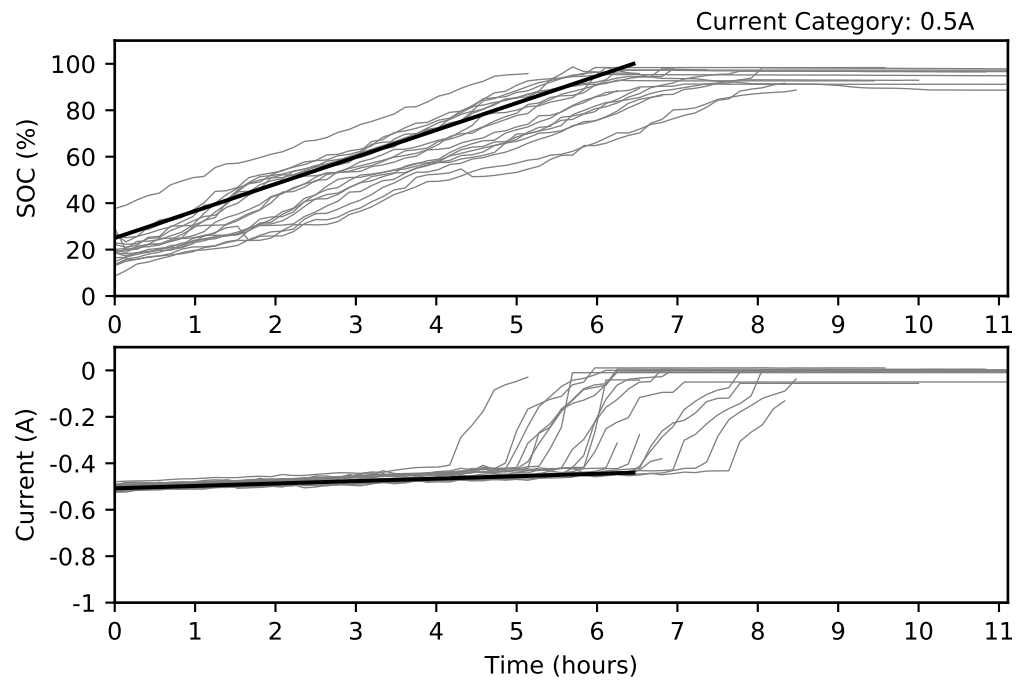
$$SOC = m_{SOC,i}t + c_{SOC,i} \quad (B.1) \quad I = m_{I,i}t + c_{I,i} \quad (B.2)$$

$$\bar{m}_{SOC} = \frac{\sum_{i=1}^n m_{SOC,i}}{n} \quad (B.3)$$

Applying this approach to the data, two different charging currents were observed, 0.7A and 0.5A, caused by changes to software of the batteries during the monitoring period. The sample of fully logged charging events was split into two and analysis for each group conducted separately. Figure B.3 shows the SOC and current plots for the fully logged charging profiles for charging currents of 0.7A and 0.5A, along with the average gradients found for each. Just 10% of the fully-logged charging events are shown for clarity. Table B.3 summarizes the average gradients and the number of fully logged charging events in each sample group. Within each current category, the charging events show very similar SOC gradients over time, with a levelling out occurring around 80%–90% SOC as expected. The charging currents are effectively constant, showing a small constant reduction over time, until the SOC reaches 80%–90%, when there is a rapid current decrease. The SOC tends to reach an asymptote around 90–95% while the battery is still connected to the charging ports; however, the SOC value was observed to jump up when the battery is disconnected and switched on again. Values higher than 1.0 have been observed at those times, showing the approximate nature of the SOC reading.



(a) Charging current 0.7A



(b) Charging current 0.5A

Figure B.3: Fully-logged charging events in grey with charging current 0.7A (B.3a) and 0.5A (B.3b), with average gradient for SOC and current in black. Just 10% of the fully-logged charging events are shown in each case for clarity.

B.4.3 Filling in Missing Charging Data

Firstly, each discontinuity was processed to determine which category of charging current it belonged to, according to the last available current reading before logging stopped. The typical charging profiles of SOC and current were then superimposed, and the last recorded voltage reading was used for the whole infill. There were two types of discontinuities, and each required a different filling strategy.

The first type was when logging had stopped during charging, the battery had been unplugged from the charging hub, and logging recommenced when it was switched on. A full charge was assumed to have occurred if the first SOC reading after the discontinuity was $\text{SOC} \geq 80\%$, because of the unreliability of the SOC measurement. If this was the case, the typical charging profile was superimposed using the appropriate gradients up to an SOC of 100% in all cases. If the first SOC reading after the discontinuity was $\text{SOC} < 80\%$, it was assumed that some issue, technical or human related, must have occurred to disrupt the charging, and the charging profile was only superimposed up to that SOC value. Although this approach may not be accurate, as it is known that the SOC values recorded can be unreliable, it was the best option given the available information.

The second type of discontinuity occurred when logging stopped during charging, but restarted while the battery was still connected to the charging hub and charging. This likely occurred if charging was disrupted so/or the battery exited charging mode (it is not known why this might have occurred, but is expected to be an internal issue with the battery), then restarted data logging, before re-entering charging mode as it detected the charging current again. When the logging starts again in these situations, the SOC readings are volatile, starting higher than expected and falling by about 0.2 over the following 30 seconds. For this reason, rather than using a linear interpolation between the SOC reading before and after the discontinuity, which would give an unrealistically high SOC gradient, the typical charging gradients are used to fill in the whole gap between the readings.

B.5 Power and Duration of Use Limits

The battery profiles showed data for some days was not representative of actual use. For example, batteries were left on without powering a load, were left at the charging hub fully charged and connected, or flicked off and on for a couple of minutes, perhaps to check the

SOC of the battery or during repair. It was important that the power data represent actual use, rather than times when the battery was inactive or being repaired at the charging hub. Therefore, power limits and duration of use limits were used, to remove data readings that were below a certain power value or duration.

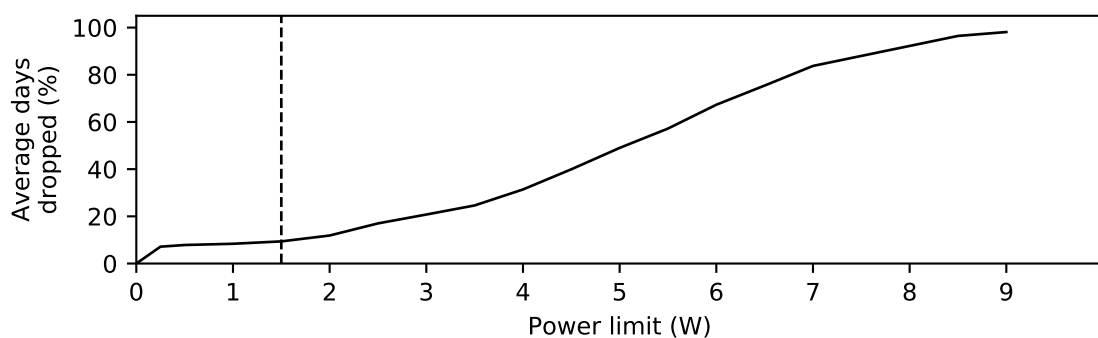
Charging and demand were investigated separately. To find the values at which the power and duration limits should be set, the percentage of use days eliminated for each profile as limits were increased was calculated. A plot of the average percentage use days eliminated was made for each limit. From this plot, the value of the power and duration limits could be inferred to remove days that do not reflect actual use, while ensuring the majority of actual use days remained.

For both demand and charging separately, and for power and duration respectively, the shape of the plot of the average percentage days eliminated with increasing value of the limit is expected to have an initial ramp up, before levelling off. This would represent elimination of the days where, for example, the batteries have been left connected at the charging hub having been fully charged. After those days have been removed, the percentage day eliminated would level out, and gradually increase. Therefore, the limit should be placed after the initial ramp up has levelled out, but before actual use days become affected, which would be shown by an increase in the gradient of the percentage days eliminated line.

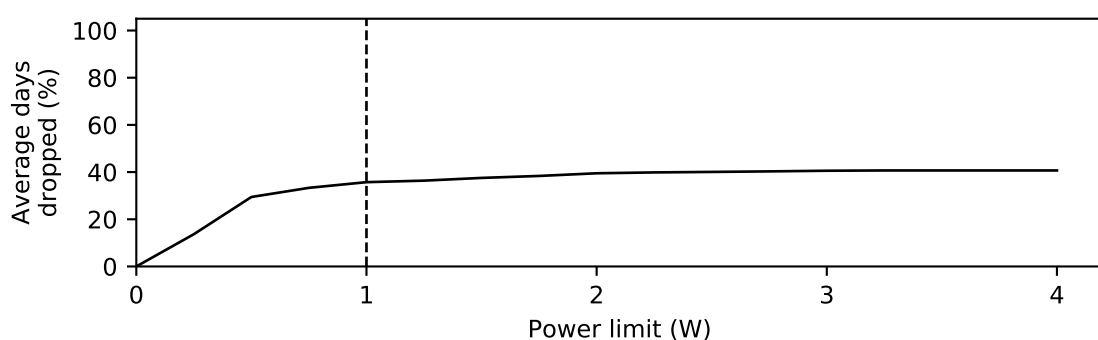
Figure B.4 shows the four plots of the average percentage days eliminated with increasing power and duration limits, for charging and demand separately.

Figure B.4a, showing the power limit applied to demand profiles, shows the expected initial ramp of days eliminated before levelling off. These are the days where parasitic power has been measured while the battery was on but not under load. The gradient then begins to increase again from 2W. This is because the minimum actual use of the system was to power one light of 3W. Around 3W more days begin to be eliminated, as they are days where only one light has been used. Therefore, the power limit is set to 1.5W, after the initial removal of days with no load, before removing days during which only one light was used.

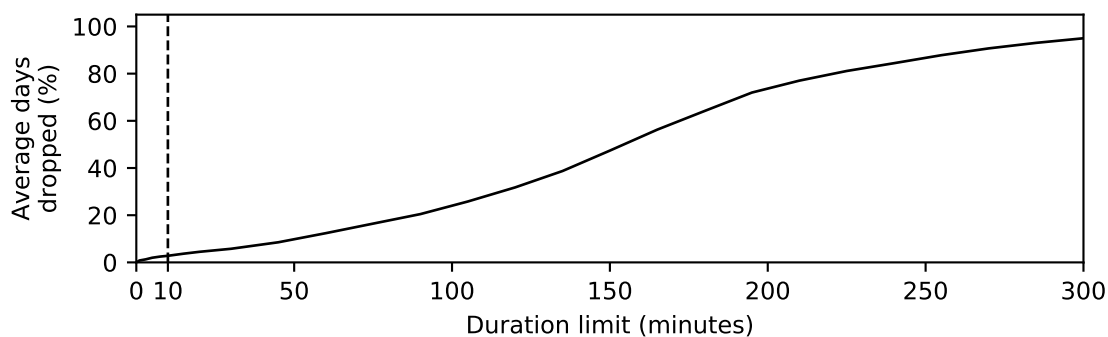
Figure B.4b, showing the equivalent power limit plot but for charging profiles, also shows the expected initial ramp, though it is more significant than the demand case. This reflects the many days where batteries were left plugged in at the hub after being repaired, and are fully charged but receiving very low trickle charge. The charging power limit is set at 1W, after the ramp has levelled off.



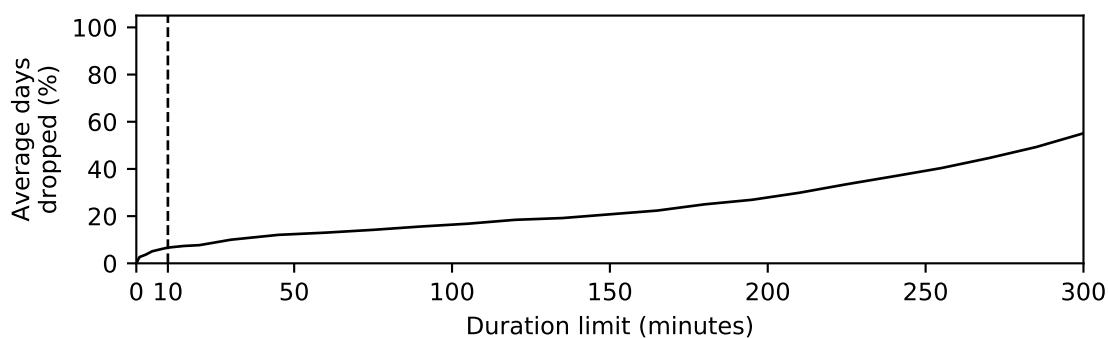
(a) Power limit for demand



(b) Power limit for charging



(c) Duration of use limit for demand



(d) Duration of use limit for charging

Figure B.4: Average percentage days dropped with increasing power and duration of use limits for demand and charging profiles.

Figure B.4c shows the duration limit applied to demand profiles. The initial ramp is very small, showing that this issue did not affect many days. Inspection of the profiles showed that when batteries were used in the morning, they were often used for short times, under 30 minutes in duration. Therefore, to avoid removing morning electricity data, the duration limit was set at 10 minutes.

Figure B.4d shows the equivalent duration limit plot for charging, and there is a clear initial ramp and subsequent levelling out where days with short charging events are eliminated. The limit is set to 10 minutes, after the ramp is levelled out.

B.6 Testing Loading Conditions to Ensure ‘Normal’ Use Data

Electricity consumption can differ between loading conditions. For example, in the UK, household energy use in the winter is different to energy use in the summer, because there are different electrical loads and typical behaviours. Weekends are often spent differently than weekdays, when many people go to school and work and so the magnitude and time of energy use differs. In addition, abnormal and uncommon loading conditions can cause abnormal energy use, such as seen in the UK during bank holidays. In a data set where electricity consumption will be compared between users, if different loading conditions correspond to different electricity use, it is important that common loading conditions are isolated and used for the analysis. Identifying and removing abnormal data that correspond to abnormal loading conditions is also important, as the objective is to analyze typical electricity use for each household.

Whether or not particular loading conditions correspond to different electricity use depends on the culture and practices of the users. Categories of loading conditions that might be important are: days of the week, public holidays, weekends, religious days, seasons and associated weather patterns. To investigate loading conditions for this electricity data, the data were compared across months, and by day of the week. Monthly comparisons were used to assess whether seasons or weather patterns affected electricity use within the monitoring period, and day of the week comparison was used to assess whether weekdays were different from weekends. Two methods were used to compare loading condition by month and by day:

1. Representative load profiles (RLPs) for each loading condition: RLPs were calculated for each loading condition by averaging over all candidate days. These were compared to each other visually to see if particular loading conditions are associated with different electricity use behaviour.
2. Comparison of electricity used per day: The electricity used each day, averaged across households, was tracked and plotted. This enables comparison across days and over time and is useful for identifying the effect of periodic loading conditions on all scales, such as day of the week or seasonal.

The majority of electricity profiles span from the end of June 2016 to the end of November 2016. One household has data extending to January 2017.

B.6.1 Monthly Electricity Use Variation

Changes in electricity use over months may be caused by changes in external temperature, weather patterns, and social factors. However, given Kenya's geographical position near the equator, only some of these factors are relevant.

Given Kenya is on the equator, it has a relatively small yearly fluctuation in temperature. In Nairobi in 2016, minimum, average and maximum-recorded temperatures fluctuated by ± 2 degrees from June to November [150]. In addition, being on the equator means that daylight hours and time of sunrise and sunset have very low variation across the year, changing by ± 20 minutes [49]. Precipitation, on the other hand, does vary throughout the year in Kenya due to two rainy seasons. The first and main rainy season occurs from March to June, and there is a shorter rainy season in November to December [150]. The monitoring period misses the first and heaviest rainy season, and coincides with the start of the short rainy season. Therefore, it is possible that household electricity use might change in November, as a result of the short rainy season, if this caused differences in environment and behaviour that led to different electricity practices.

Social factors that may impact electricity use include the rhythms of the school year, as schools are closed in August in Kenya. This might change daily household routines, reflected in the electricity use patterns of households.

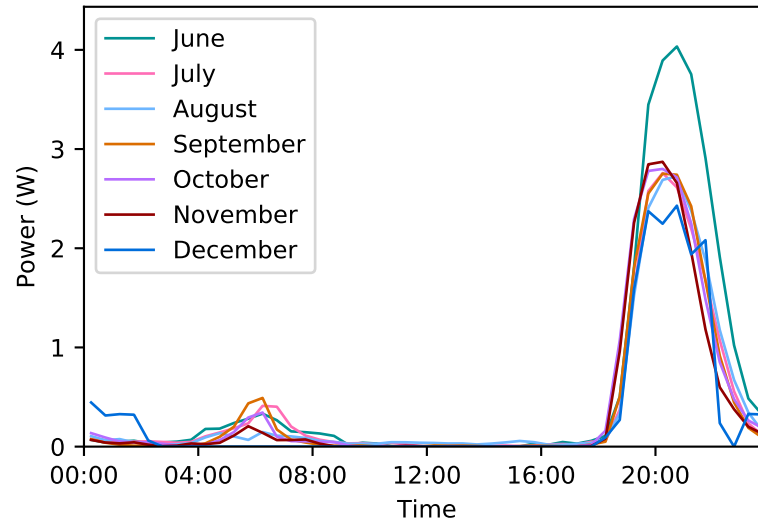


Figure B.5: Representative load profiles for each month, calculated by averaging all days for all households for each month.

Therefore, August had the potential to cause different electricity use due to school closures, and November due to the small rainy seasons. Weekends were removed for the analysis just in case they had very different electricity patterns, given they had not been tested yet.

Figure B.5 shows the monthly RLPs, where evening use in June is much higher than in all other months, which are very similar to each other. Figure B.6 shows the average electricity consumption per weekday for all days for which there is electricity data. This calculation only included households that actually used electricity on that day, ignoring those who did not happen to use their household battery that day. The analysis was repeated splitting the day into two, from 00:00-12:00, and 12:01-00:00, to see if either morning or evening use were affected separately, but the results are the same as the total daily evening plot. The plot shows the days at the end of June, when the project started, had higher electricity use than later in the monitoring period. This initial high use could have occurred because at the start of the project households were excited to try the system out, so were testing the capacity of the household battery before settling into a regular use pattern. Given this, June data were removed from the data sets.

Fig B.5 shows the RLP for December is less smooth than the others, but follows roughly the same trajectory. Fig B.6 shows that in December the average electricity used is erratic and much less consistent with the values from August to November. This is because there is

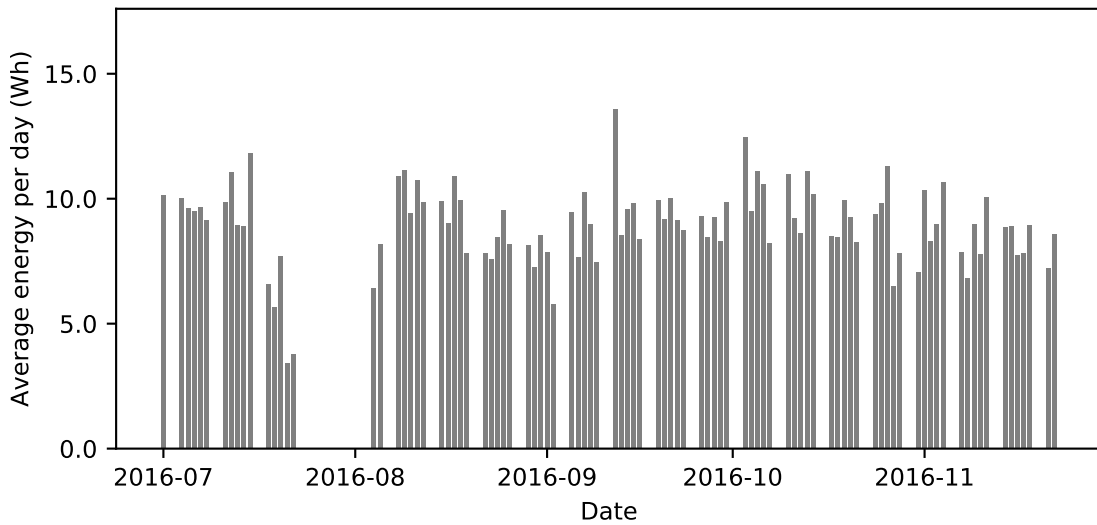


Figure B.6: Average energy used each day across all households during the monitoring period.

data for just one household after the 21st November, and only 16 days contribute to the RLP. Therefore, data readings from the 22nd November inclusive were removed.

The analysis shows that electricity use was not affected, as hypothesized, by the short rainy season in November or school closures in August. Instead, electricity use was different at the beginning of the project, as households were getting used to the system. Therefore, all the data readings from 1st July to 21st November inclusive were retained.

B.6.2 Day of the Week Electricity Use Variation

Electricity might be used differently on different days of the week if there are different activities that occur according to the day. According to local sources, activities on Saturdays can be the same as weekdays for small business owners and farmers in rural Kenyan communities such as Lomolo B and Echariria, because they go about their business or farm as usual. However, Sundays might be different because religious activities occur.

The analysis on days of the week contained only days from the 1st July to 21st November inclusive. Figure B.7 shows the RLPs for each day of the week, and Figure B.8 shows the average electricity used per day, with weekdays and weekends coloured differently. The RLPs for each day of the week show that in the evenings, electricity use on Saturdays and Sundays is within the variance of the other weekdays. However in the morning, use on Saturdays and Sundays is slightly lower than the electricity use on other days, though the magnitude of average power difference is only in order of 0.2W, which is very small. The average

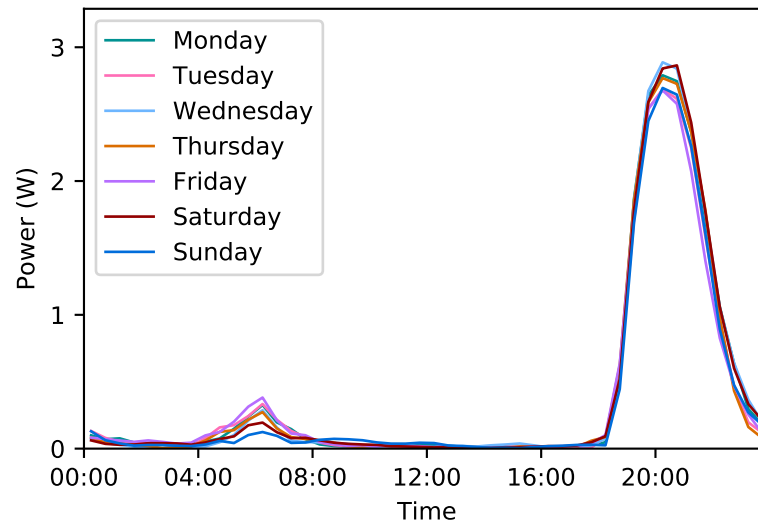


Figure B.7: Representative load profiles for each day of the week, calculated by averaging all days for all households for each day of the week, between 1st July and 21st November inclusive.

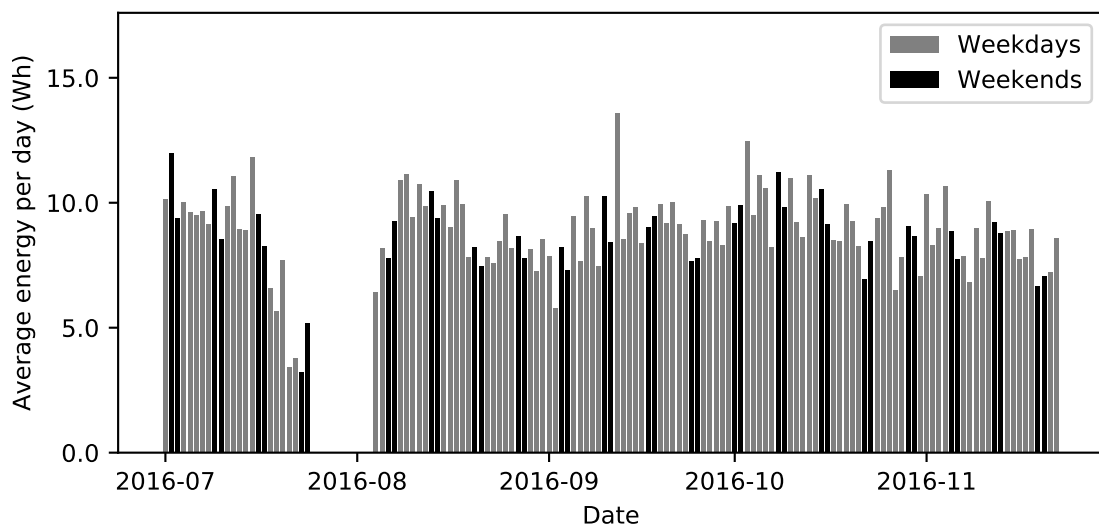


Figure B.8: Average energy used each day across all households for weekdays (grey) and weekend days (black).

daily electricity consumption plot does not show a discernible difference in electricity use between weekdays and weekend days. Despite the small difference in morning electricity use between weekdays and weekend days, all days of the week were retained in the data set because existing differences were very small. If analysis were to be carried out solely on morning electricity use, removal of Sundays and possibly Saturdays could be considered to ensure analysis of morning use was not affected by that variation.

C

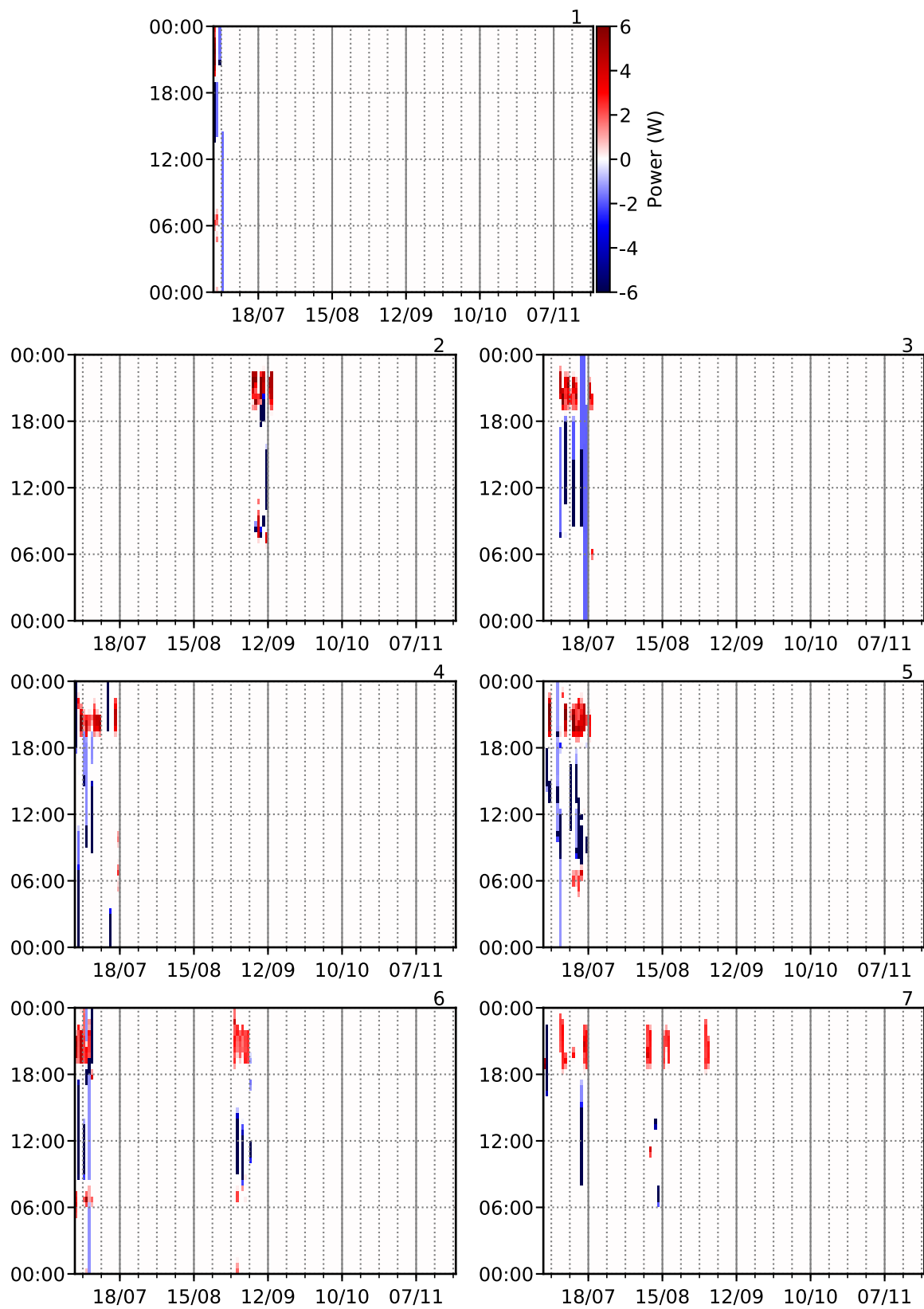
Electricity Data

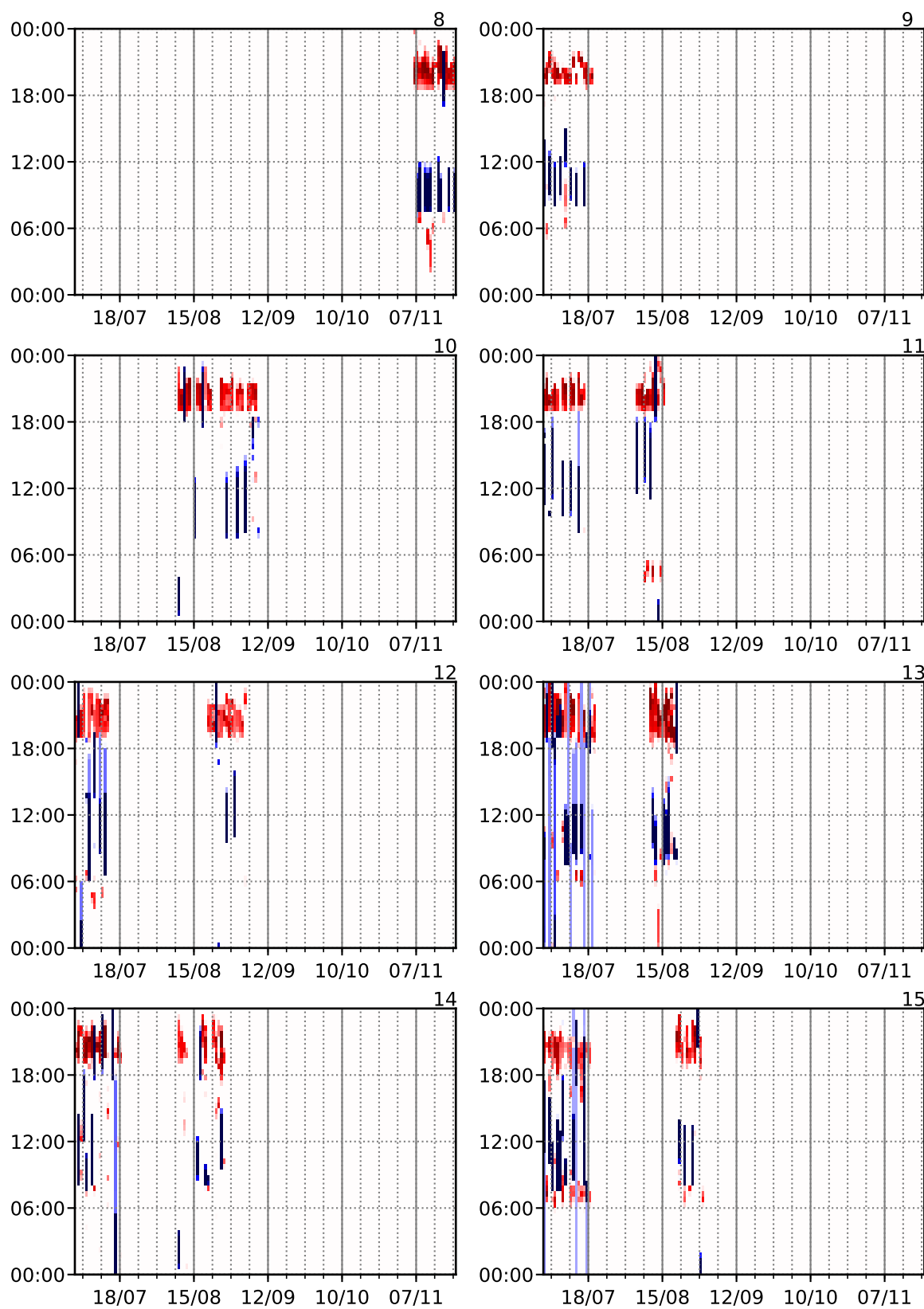
Figure C.1 shows the 51 household lighting power profiles as heat maps, showing demand and charging power over the whole monitoring period (regardless of the amount of data available).

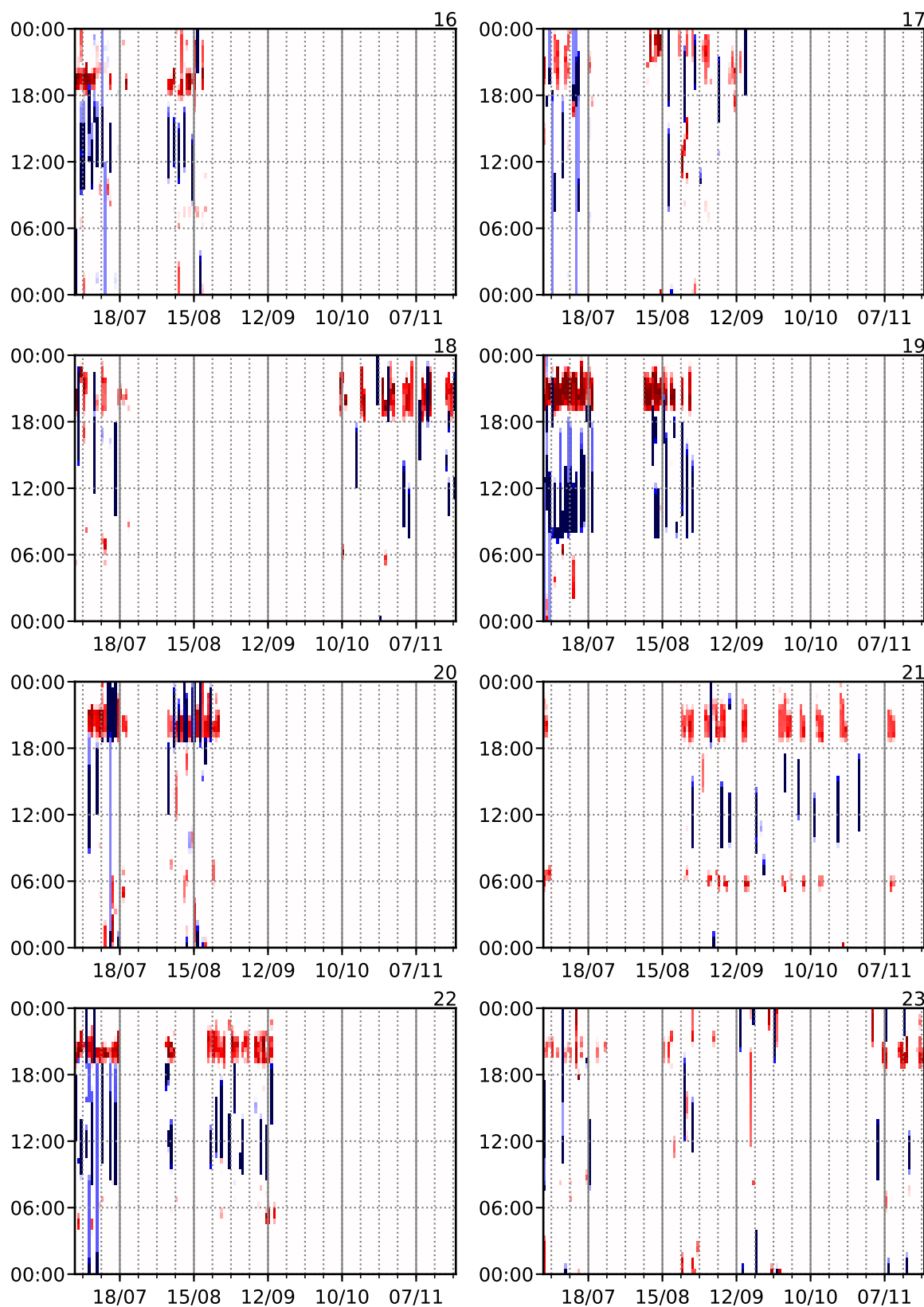
The full data set is available from Mendeley Data through the doi noted in the Preface of this thesis.

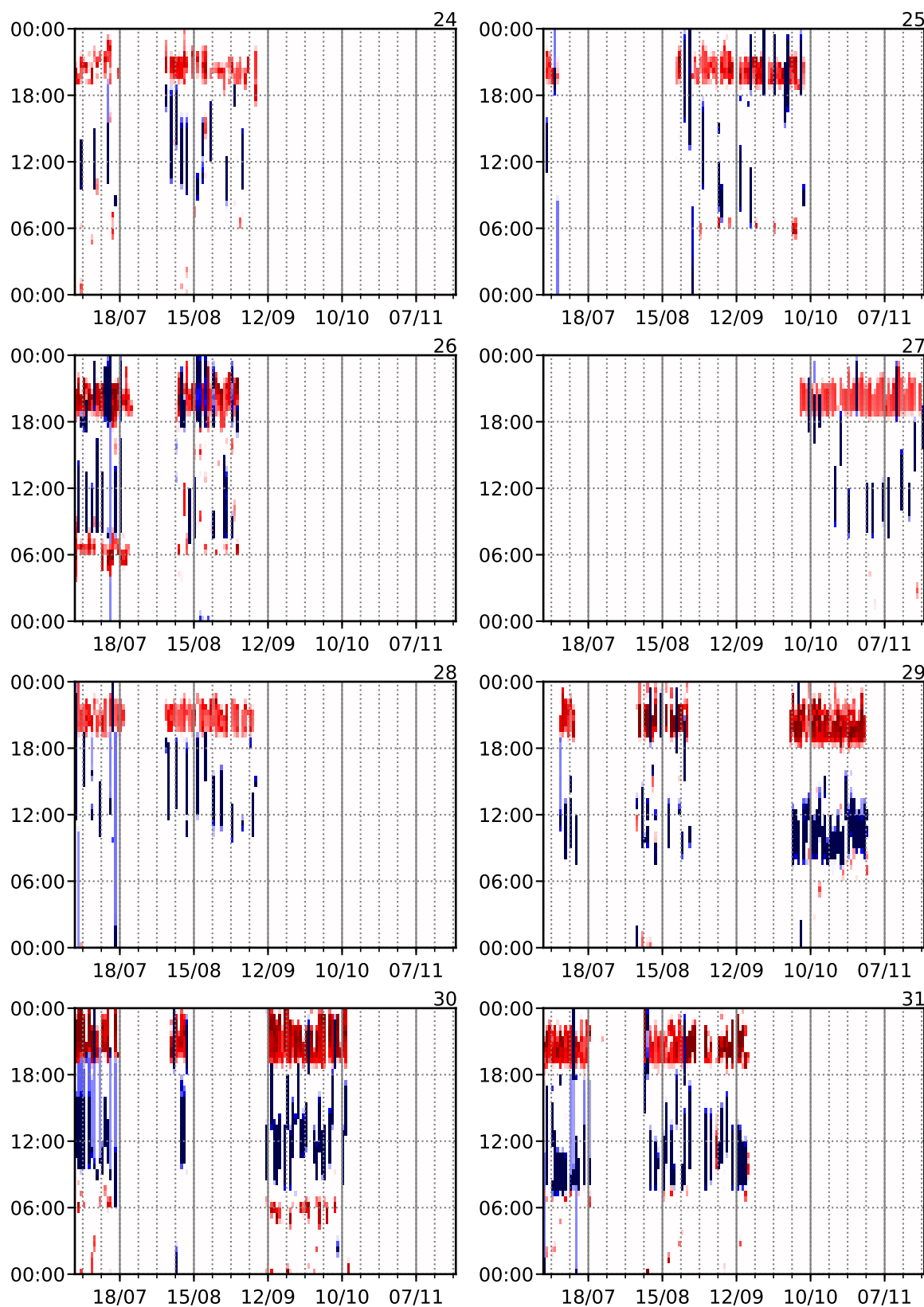
Four batteries had sections of 1970 data that were calibrated to 2016 using best estimates, as explained in Appendix B Section B.2. During the combination of user profiles step, one of the sections of data was discarded. Therefore, three profiles have sections of data which are time calibrated by estimations. These are listed here, alongside the dates concerned, and whether the 1970 calibrated data section is the sole use pattern or is combined with another:

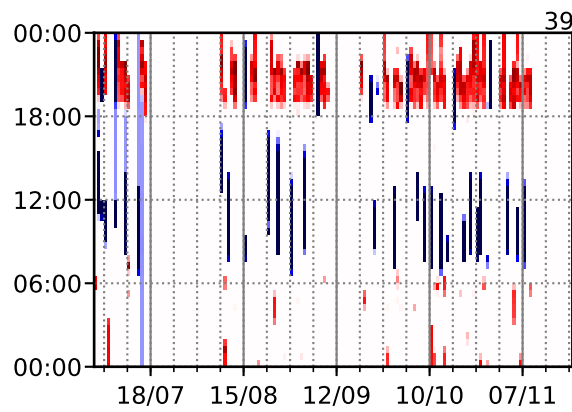
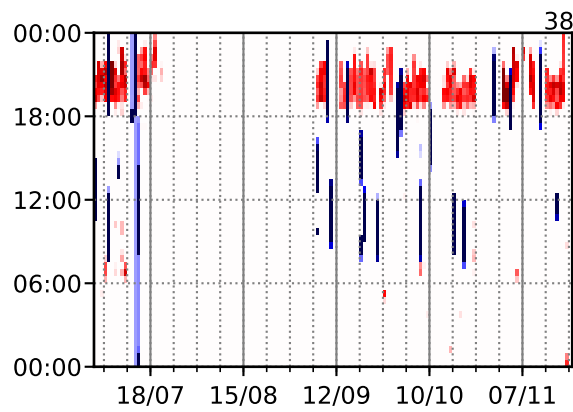
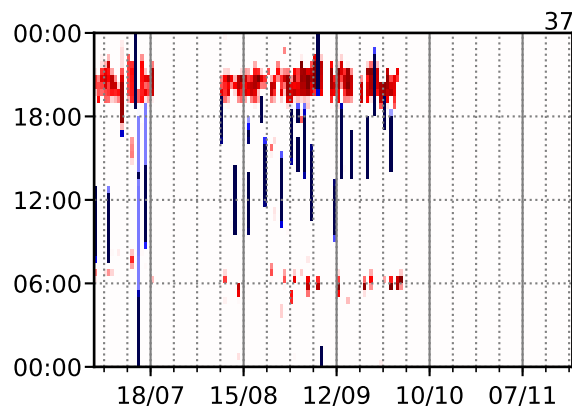
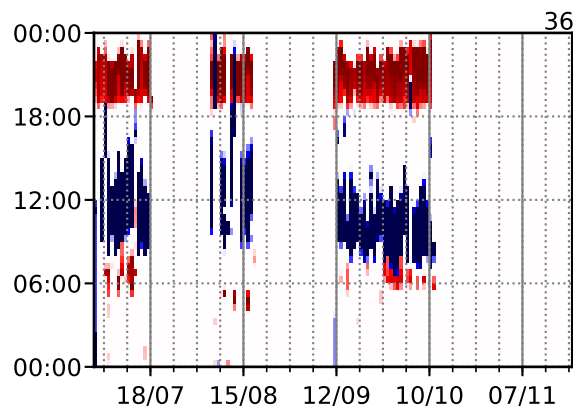
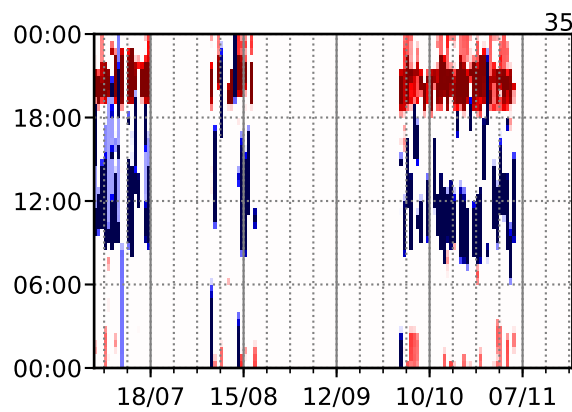
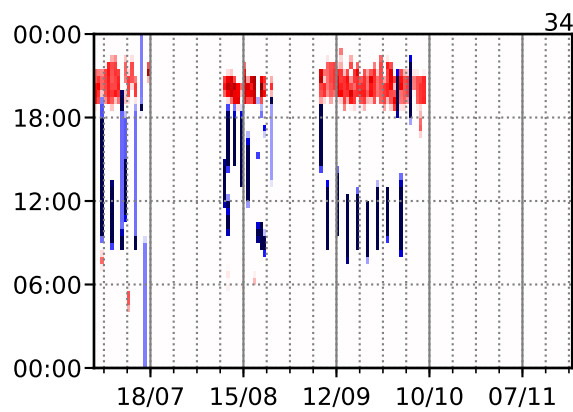
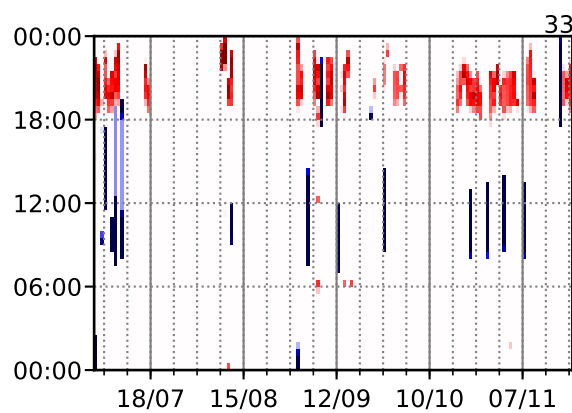
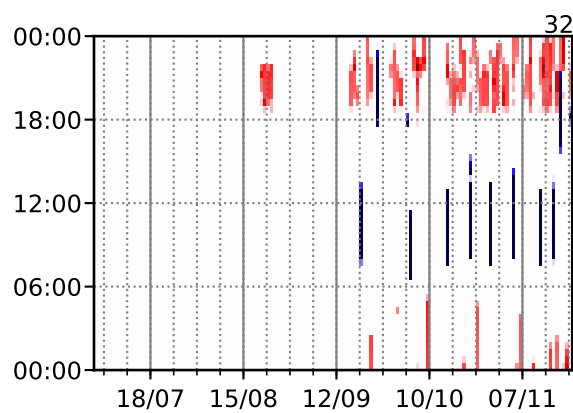
1. Profile 25: 25/08 - 06/09/16, as the sole use pattern.
2. Profile 41: 19/09 - 15/11/16, alongside a second use pattern from 19 - 28/09/16, thereafter as the sole use pattern.
3. Profile 50: 26/09 - 25/10/16, alongside a second use pattern for the duration.

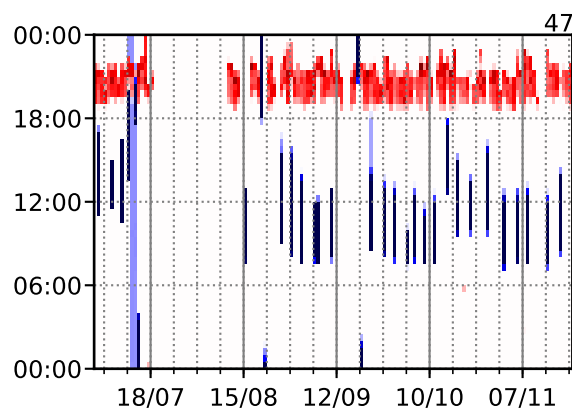
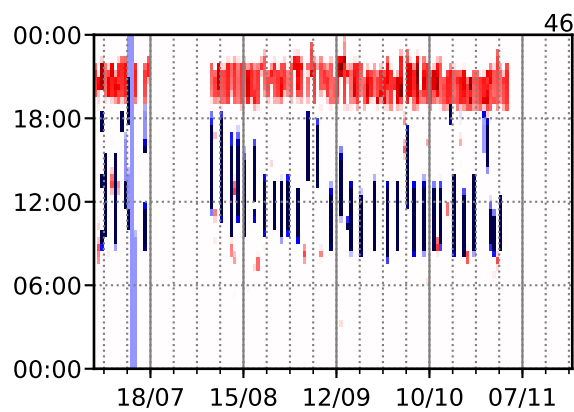
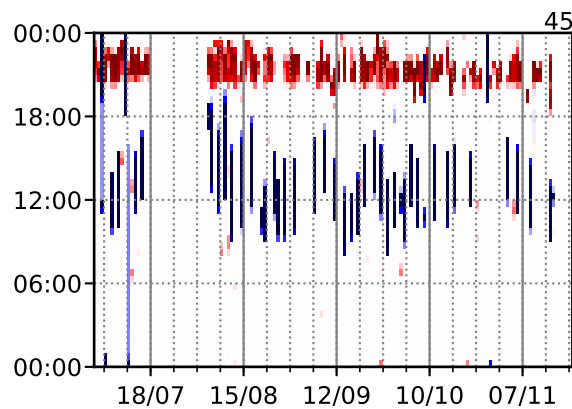
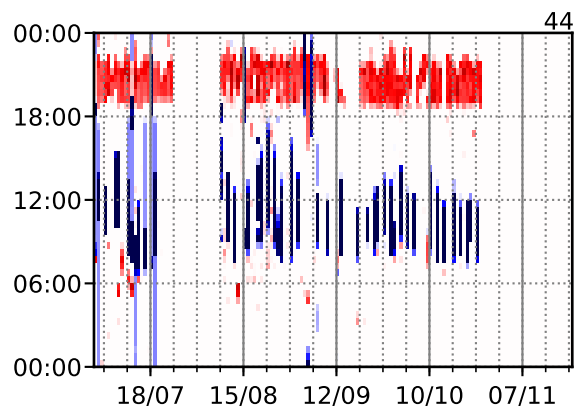
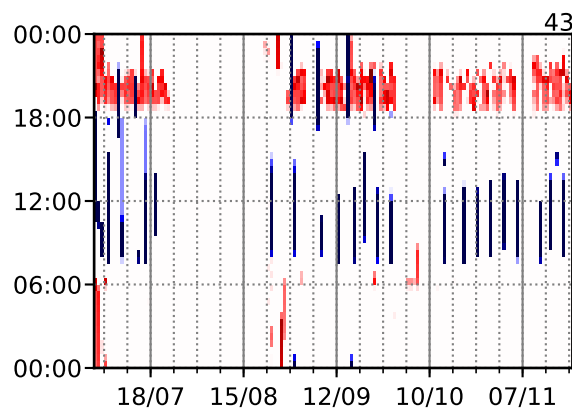
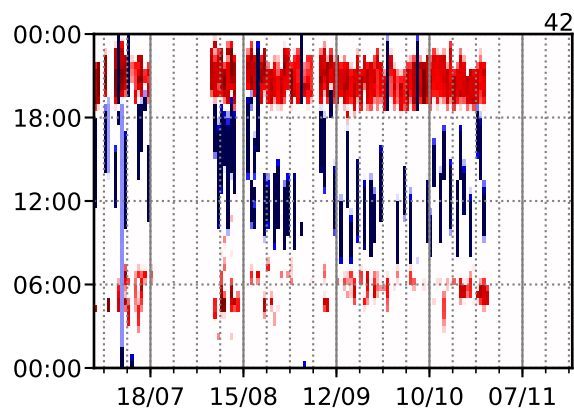
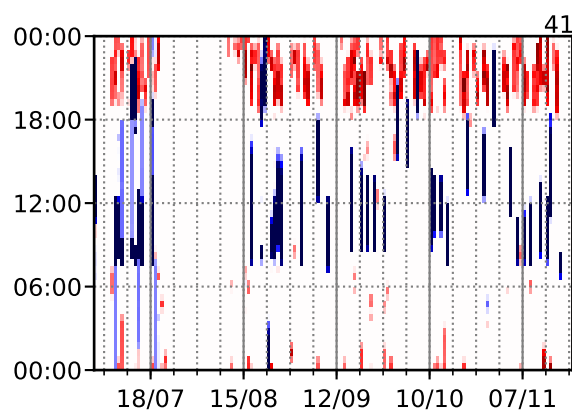
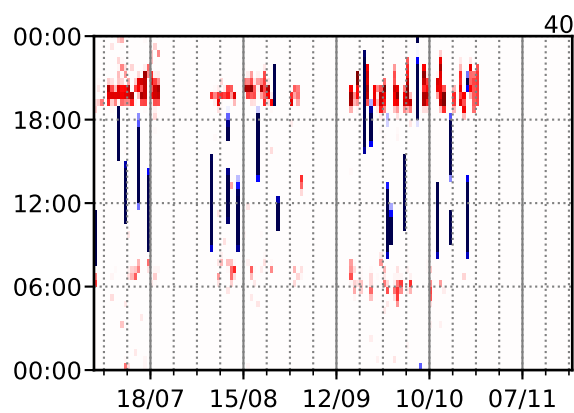












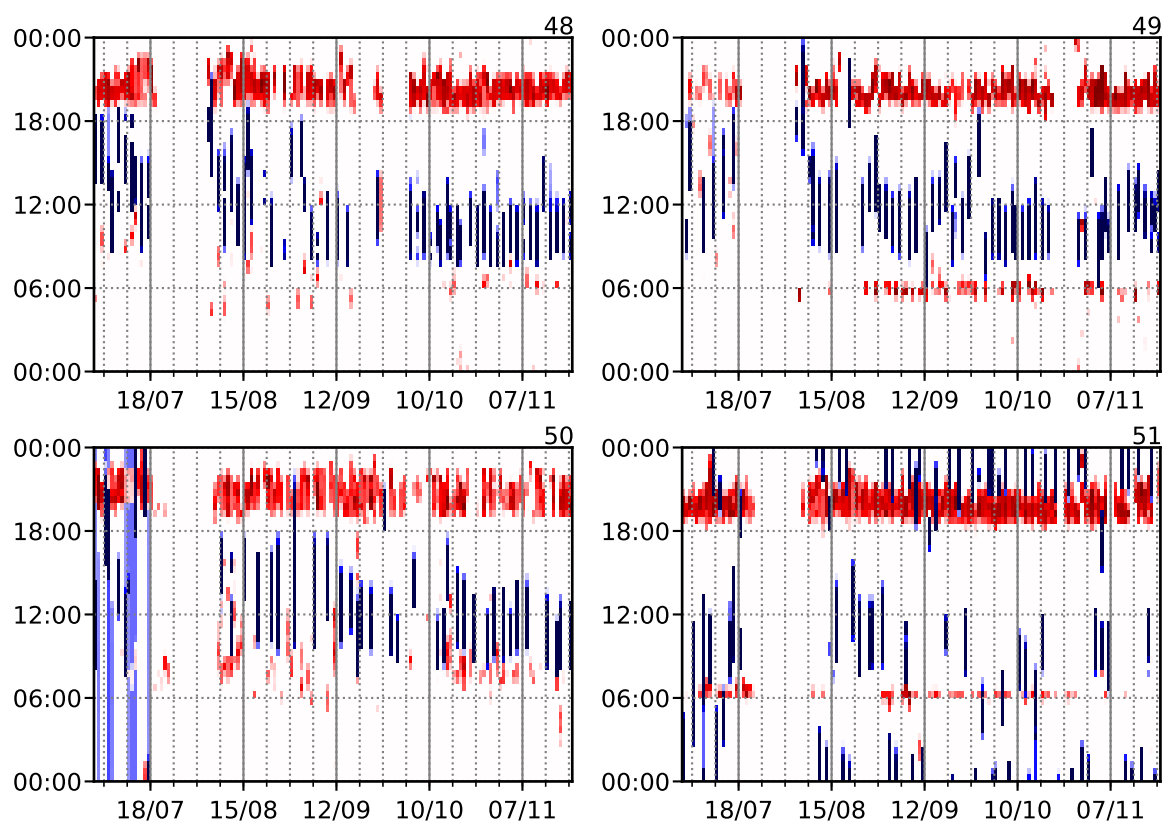


Figure C.1: 51 household power heat maps with half-hourly time resolution showing battery power on the 12V port for monitoring period.

References

- [1] International Energy Agency, “WEO Energy Access Outlook 2017: Methodology for Energy Access Analysis”, Tech. Rep., 2017.
- [2] *World Bank Open Data*. [Online]. Available: <https://data.worldbank.org>.
- [3] A. Castañeda, D. Doan, D. Newhouse, M. C. Nguyen, H. Uematsu, and J. P. Azevedo, “Who Are the Poor in the Developing World?”, World Bank Group, Tech. Rep. 7844, 2016, p. 41.
- [4] “Transforming our world: The 2030 agenda for sustainable development”, United Nations, Tech. Rep., 2015. arXiv: 1011.1669v3.
- [5] International Energy Agency, “Energy poverty: How to make modern energy access universal?”, Paris, Tech. Rep. September, 2010.
- [6] F. Fuso Nerini, J. Tomei, L. S. To, I. Bisaga, P. Parikh, M. Black, A. Borrion, C. Spataru, V. Castán Broto, G. Anandarajah, B. Milligan, and Y. Mulugetta, “Mapping synergies and trade-offs between energy and the Sustainable Development Goals”, *Nature Energy*, vol. 3, no. 1, pp. 10–15, 2018.
- [7] M. Willcox, L. Waters, H. Wanjiru, A. Pueyo, Debajit Palit, and K. R. Sharma, “Utilising Electricity Access for Poverty Reduction: Executive Summary”, Practical Action Consulting, Tech. Rep., 2015.
- [8] C. Kirubi, A. Jacobson, D. Kammen, and A. Mills, “Community-based electric micro-grids can contribute to rural development: evidence from Kenya.”, *World Development*, vol. 37, pp. 1208–1221, 2009.
- [9] O. Stojanovski, M. Thurber, and F. Wolak, “Rural energy access through solar home systems: Use patterns and opportunities for improvement”, *Energy for Sustainable Development*, vol. 37, pp. 33–50, 2017.
- [10] M. Aklin, P. Bayer, S. P. Harish, and J. Urpelainen, “Does basic energy access generate socioeconomic benefits? A field experiment with off-grid solar power in India”, *Science Advances*, vol. 3, no. 5, pp. 1–9, 2017.
- [11] N. L. Lam, K. R. Smith, A. Gauthier, and M. N. Bates, “Kerosene: A Review of Household Uses and their Hazards in Low- and Middle-Income Countries”, *Journal of Toxicology and Environmental Health, Part B*, vol. 15, no. 6, pp. 396–432, 2012.
- [12] S. Mahapatra, H. N. Chanakya, and S. Dasappa, “Evaluation of various energy devices for domestic lighting in India: Technology, economics and CO2 emissions”, *Energy for Sustainable Development*, vol. 13, no. 4, pp. 271–279, 2009.
- [13] IPCC, “Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]”, IPCC, Geneva, Tech. Rep., 2014, p. 151.
- [14] N. L. Lam, Y. Chen, C. Weyant, C. Venkataraman, P. Sadavarte, M. A. Johnson, K. R. Smith, B. T. Brem, J. Arineitwe, J. E. Ellis, and T. C. Bond, “Household light makes global heat: High black carbon emissions from kerosene wick lamps”, *Environmental Science and Technology*, vol. 46, no. 24, pp. 13 531–13 538, 2012.

- [15] C. Casillas and D. Kammen, “The Energy-Poverty-Climate Nexus”, *Science*, vol. 330, no. 6008, pp. 1181–1182, 2010.
- [16] International Energy Agency, “World Energy Outlook”, Tech. Rep., 2018.
- [17] Practical Action, “Poor People’s Energy Outlook 2016: National energy access planning from the bottom up”, Practical Action, Tech. Rep., 2016, p. 90.
- [18] M. Bhatia and N. Angelou, “Beyond Connections: Energy Access Redefined. ESMAP Technical Report 008/15”, Energy Sector Management Assistance Program. World Bank Group, Washington, DC, Tech. Rep., 2015, pp. 1–244.
- [19] Practical Action, “Poor people’s energy outlook 2014: Key messages on energy for poverty alleviation”, Practical Action Publishing, Rugby, UK, Tech. Rep., 2014.
- [20] Sustainable Energy for All, “Progress Toward Sustainable Energy: Global Tracking Framework 2015 Summary Report”, ESMAP, World Bank, IEA, Tech. Rep., 2015, p. 332.
- [21] M. Bhatia and N. Angelou, “Beyond Connections: Energy Access Redefined. Executive Summary”, Energy Sector Management Assistance Program. World Bank Group, Washington, DC, Tech. Rep., 2015.
- [22] International Energy Agency, “Energy Access Outlook 2017: From poverty to prosperity”, Tech. Rep., 2017, p. 144.
- [23] P. Alstone, D. Gershenson, and D. M. Kammen, “Decentralized energy systems for clean electricity access”, *Nature Climate Change*, vol. 5, no. 4, pp. 305–314, 2015.
- [24] P. Kemeny, P. Munro, N. Schiavone, G. van der Horst, and S. Willans, “Community Charging Stations in rural sub-Saharan Africa: Commercial success, positive externalities, and growing supply chains”, *Energy for Sustainable Development*, vol. 23, pp. 228–236, 2014.
- [25] K. Ulsrud, T. Winther, D. Palit, and H. Rohrer, “Village-level solar power in Africa: Accelerating access to electricity services through a socio-technical design in Kenya”, *Energy Research & Social Science*, 2015.
- [26] A. K. Gerlak and M. Wilder, “Exploring the textured landscape of water insecurity and the human right to water”, *Environment: Science and Policy for Sustainable Development*, vol. 54, no. 2, pp. 4–17, 2012.
- [27] UNICEF, “An evaluation of the PlayPump water system as an appropriate technology for water, sanitation and hygiene programmes”, UNICEF, New York, Tech. Rep., 2007, pp. 1–17.
- [28] A. L. Obiols and K. Erpf, “Mission Report on the Evaluation of the PlayPumps installed in Mozambique”, Mozambique: Centro de Formação Profissional de Água e Saneamento, Swiss Resource Centre, and Consultancies for Development (SKAT), Tech. Rep., 2008.
- [29] *Roundabout Water Solutions: How it works - Playpumps*. [Online]. Available: <http://www.playpumps.co.za/index.php/how-it-works/> (visited on 11/24/2018).
- [30] M. Hobbes, “Stop Trying to Save the World: Big Ideas are Destroying International Development”, *The New Republic Online*, Nov. 2014.
- [31] R. P. Byrne, “Learning drivers: rural electrification regime building in Kenya and Tanzania”, PhD Thesis, University of Sussex, 2011, p. 288.
- [32] V. G. García and M. M. Bartolomé, “Rural electrification systems based on renewable energy: The social dimensions of an innovative technology”, *Technology in Society*, vol. 32, no. 4, pp. 303–311, 2010.
- [33] M. Schäfer, N. Kebir, and K. Neumann, “Research needs for meeting the challenge of decentralized energy supply in developing countries”, *Energy for Sustainable Development*, vol. 15, no. 3, pp. 324–329, 2011.

- [34] A. Maunder, G. Marsden, D. Gruijters, and E. Blake, “Designing interactive systems for the developing world - reflections on user-centred design”, in *International Conference on Information and Communication Technologies and Development*, 2007, pp. 1–8.
- [35] K. M. Bryden and N. G. Johnson, “Understanding rural village energy needs and design constraints”, in *Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, 2011, pp. 1–10.
- [36] IEEE, *IEEE Standard 1562: IEEE guide for array and battery sizing in stand-alone photovoltaic (PV) systems*, New York, 2008.
- [37] N. Agarwal, A. Kumar, and Varun, “Optimization of grid independent hybrid PV-diesel-battery system for power generation in remote villages of Uttar Pradesh, India”, *Energy for Sustainable Development*, vol. 17, no. 3, pp. 210–219, 2013.
- [38] R. Sen and S. C. Bhattacharyya, “Off-grid electricity generation with renewable energy technologies in India: An application of HOMER”, *Renewable Energy*, vol. 62, pp. 388–398, 2014.
- [39] H. Louie and P. Dauenhauer, “Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability”, *Energy for Sustainable Development*, vol. 34, pp. 30–43, 2016.
- [40] P. Boait, V. Advani, and R. Gammon, “Estimation of demand diversity and daily demand profile for off-grid electrification in developing countries”, *Energy for Sustainable Development*, vol. 29, pp. 135–141, 2015.
- [41] P. Siano, “Demand response and smart grids - A survey”, *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.
- [42] SESDC, *SESDC Energy Use Archive*, IEEE PES Sustainable Energy Systems for Developing Communities Working Group Energy Use Archive, 2017. [Online]. Available: <http://sites.ieee.org/pes-sesdc/output/data-sets/energy-use-archive/>.
- [43] ‘Home: Peak’, 2018. [Online]. Available: <https://www.powerforall.org/peak/> (visited on 08/30/2018).
- [44] C. Blodgett, P. Dauenhauer, H. Louie, and L. Kickham, “Accuracy of energy-use surveys in predicting rural mini-grid user consumption”, *Energy for Sustainable Development*, vol. 41, pp. 88–105, 2017.
- [45] *Global LEAP Awards*. [Online]. Available: <https://globalleapawards.org> (visited on 05/19/2019).
- [46] N. J. Williams, P. Jaramillo, B. Cornell, I. Lyons-Galante, and E. Wynn, “Load Characteristics of East African Microgrids”, in *IEEE PES-IAS PowerAfrica*, IEEE, 2017, pp. 236–241.
- [47] J. Stephenson, B. Barton, G. Carrington, A. Doering, R. Ford, D. Hopkins, R. Lawson, A. McCarthy, D. Rees, M. Scott, P. Thorsnes, S. Walton, J. Williams, and B. Wooliscroft, “The energy cultures framework: Exploring the role of norms, practices and material culture in shaping energy behaviour in New Zealand”, *Energy Research and Social Science*, vol. 7, pp. 117–123, 2015.
- [48] M. R. Khan and E. D. Brown, “DC nanogrids: A low cost PV based solution for livelihood enhancement for rural Bangladesh”, in *Innovating Energy Access for Remote Areas: Discovering Untapped Resources. Proceedings of the International Conference, 10th-12th April 2014, University of California, Berkeley*, M. Schäfer, D. Kammen, N. Kebir, and D. Philipp, Eds., Berlin: Universitätsverlag der TU Berlin, 2014, pp. 79–83.
- [49] *Daylight times data*. [Online]. Available: <https://www.timeanddate.com> (visited on 01/01/2019).

- [50] J. Karat and C. M. Karat, "The evolution of user-centered focus in the human-computer interaction field", *IBM Systems Journal*, vol. 42, no. 4, pp. 532–541, 2003.
- [51] E. B.-N. Sanders and P. J. Stappers, "Co-creation and the new landscapes of design", *CoDesign*, vol. 4, no. 1, pp. 5–18, 2008.
- [52] B. Moggridge and B. Atkinson, *Designing Interactions*, Volume 17. Cambridge, MA: MIT press, 2007.
- [53] D. Sangiorgi, "Building a framework for service design research", *8th European Academy Of Design Conference*, vol. 28, no. April, pp. 415–420, 2009.
- [54] C. Burns, H. Cottam, C. Vanstone, and J. Winhall, "Transformation Design", UK Design Council, Tech. Rep., 2006, pp. 1–33.
- [55] L. Kimbell, "Rethinking Design Thinking: Part I", *Design and Culture*, vol. 3, no. 3, pp. 285–306, 2011. arXiv: arXiv:1011.1669v3.
- [56] S. Di Russo, "Understanding the behaviour of design thinking in complex environments", PhD thesis, 2016, p. 310.
- [57] R. Buchanan, "Wicked Problems in Design Thinking", *Design Issues*, vol. 8, no. 2, pp. 5–21, 1992.
- [58] H. W. Rittel and M. M. Webber, "Dilemmas in a General Theory of Planning", *Policy Sciences*, vol. 2, pp. 155–169, 1973.
- [59] C. Watkins, G. Loudon, S. Gill, and J. Hall, "The Challenges of Taking a User-Centric Approach within Developing Countries: A Case Study of Designing Medical for Zambia", in *11th European Academy of Design Conference: The Value of Design Research*, 2015.
- [60] D. Ramachandran, M. Kam, J. Chiu, J. Canny, J. L. Frankel, and S. Monica, "Social Dynamics of Early Stage Co-Design in Developing Regions", in *CHI 2007 Proceedings: Designing for Specific Cultures*, San Jose, CA, USA, 2007, pp. 1087–1096.
- [61] T. Parikh, K. Ghosh, and A. Chavan, "Design studies for a financial management system for micro-credit groups in rural india", *ACM SIGCAPH Computers and the Physically Handicapped*, no. 73-74, pp. 15–22, 2003.
- [62] S. Hussain, E. B. N. Sanders, and M. Steinert, "Participatory Design with Marginalized People in Developing Countries: Challenges and Opportunities Experienced in a Field Study in Cambodia", *International Journal of Design*, vol. 6, no. 2, pp. 91–109, 2013.
- [63] I. Medhi, A. Sagar, and K. Toyama, "User Interfaces for Non-Literate and Semi-Literate Users", *Information and Communication Technologies and Development, 2006. ICTD'06*, vol. 4, no. 1, pp. 72–82, 2006.
- [64] M. Chetty and R. E. Grinter, "HCI4D: HCI Challenges In The Global South", in *Computer-Human Interaction (CHI) '07 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2007, pp. 2327–2332.
- [65] C. Putnam, E. Rose, and E. J. Johnson, "Adapting User-Centered Design Methods to Design for Diverse Populations", *Information Technologies and International Development*, vol. 5, no. 4, pp. 51–73, 2009.
- [66] A. Mohr, "'Opening up' energy transitions research for development", in *Science and the Politics of Openness: Here be Monsters*, B. Nerlich, S. Hartley, S. Raman, and A. Smith, Eds., Manchester University Press, 2018, pp. 1–16.
- [67] J. Cloke, A. Mohr, and E. Brown, "Imagining renewable energy: Towards a Social Energy Systems approach to community renewable energy projects in the Global South", *Energy Research and Social Science*, vol. 31, no. October 2016, pp. 263–272, 2017.

- [68] IDEO.org, *The Field Guide to Human-Centered Design*, 1st. 2015.
- [69] S. Hirmer and P. Guthrie, “Identifying the needs of communities in rural Uganda: A method for determining the ‘User-Perceived Value’ of rural electrification initiatives”, *Renewable and Sustainable Energy Reviews*, vol. 66, pp. 476–486, 2016.
- [70] A. Cooper, R. Reimann, and D. Cronin, *About Face 3.0: The Essentials of Interaction Design*. John Wiley & Sons, Inc., 2007.
- [71] S. Bødker and E. Christiansen, “Scenarios as springboards in design”, *Social Science Research, Technical Systems and Cooperative Work*, pp. 217–234, 2003.
- [72] C. N. Chapman, E. Love, R. P. Milham, P. Elrif, and J. L. Alford, “Quantitative Evaluation of Personas as Information”, in *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, New York., 2008, pp. 1107–1111.
- [73] E. Beck, M. Obrist, R. Bernhaupt, M. Tscheligi, and T. Cedex, “Instant Card Technique : How and Why to Apply in User-Centered Design”, in *Proceedings of the Tenth Anniversary Conference on Participatory Design*, 2008, pp. 162–165.
- [74] S. C. Weller and A. K. Romney, *Systematic data collection*. Newbury Park, Calif.: Sage Publications, 1988.
- [75] H. Beyer and K. Holtzblatt, *Contextual design: defining customer-centered systems*. Elsevier, 1997.
- [76] R. Pereira and M. C. C. Baranauskas, “A value-oriented and culturally informed approach to the design of interactive systems”, *Journal of Human Computer Studies*, vol. 80, pp. 66–82, 2015.
- [77] O. Mont, “Clarifying the concept of product – service system”, *Journal of Cleaner Production*, vol. 10, pp. 237–245, 2002.
- [78] S. Emili, F. Ceschin, and D. Harrison, “Product-Service System applied to Distributed Renewable Energy: A classification system, 15 archetypal models and a strategic design tool”, *Energy for Sustainable Development*, vol. 32, pp. 71–98, 2016.
- [79] C. Vezzoli, C. Kohtala, A. Srinivasan, L. Xin, M. Fusakul, D. Sateesh, and J. C. Diehl, *Product-Service System Design for Sustainability*. London: Greenleaf, 2014.
- [80] C. Vezzoli, F. Ceschin, L. Osanjo, M. K. M’Rithaa, R. Moalosi, V. Nakazibwe, and J. C. Diehl, *Designing Sustainable Energy for All: Sustainable Product-Service System Design Applied to Distributed Renewable Energy*. Springer, 2018.
- [81] P. Sandwell, C. Chambon, A. Saraogi, A. Chabenat, M. Mazur, N. Ekins-Daukes, and J. Nelson, “Analysis of energy access and impact of modern energy sources in unelectrified villages in Uttar Pradesh”, *Energy for Sustainable Development*, vol. 35, pp. 67–79, 2016.
- [82] N. U. Blum, R. Sryantoro Wakeling, and T. S. Schmidt, “Rural electrification through village grids - Assessing the cost competitiveness of isolated renewable energy technologies in Indonesia”, *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 482–496, 2013.
- [83] D. Fabini, D. Ponce, D. L. Baridó, A. Omu, and J. Taneja, “Mapping Induced Residential Demand for Electricity in Kenya”, in *ACM DEV-5 ’14 Proceedings of the Fifth ACM Symposium on Computing for Development*, vol. 5, 2014, pp. 43–52.
- [84] S. Mandelli, M. Merlo, and E. Colombo, “Novel procedure to formulate load profiles for off-grid rural areas”, *Energy for Sustainable Development*, vol. 31, pp. 130–142, 2016.
- [85] N. Narayan, Z. Qin, J.-C. Diehl, J. Popovic-Gerber, P. Bauer, and M. Zeman, “Stochastic load profile construction for the multi-tier framework for household electricity access using off-grid DC appliances”, *Energy Efficiency*, 2018.

- [86] E. Hartvigsson and E. O. Ahlgren, "Comparison of load profiles in a mini-grid: Assessment of performance metrics using measured and interview-based data", *Energy for Sustainable Development*, vol. 43, pp. 186–195, 2018.
- [87] I. Bisaga, N. Puźniak-Holford, A. Grealish, C. Baker-Brian, and P. Parikh, "Scalable off-grid energy services enabled by IoT: A case study of BBOX SMART Solar", *Energy Policy*, vol. 109, no. July, pp. 199–207, 2017.
- [88] K. Weber, S. Kennedy, and P. Armstrong, "End-Use Load Monitoring of a Micro Hydroelectric-Powered Community Micro Grid: A Case Study in Rural Malaysia", in *Innovating Energy Access for Remote Areas: Discovering Untapped Resources. Proceedings of the International Conference, 10th-12th April 2014, University of California, Berkeley*, M. Schäfer, D. Kammen, N. Kebir, and D. Philipp, Eds., Berlin: Universitätsverlag der TU Berlin, 2014.
- [89] J. Deign, *What Smart Meters Tell Us About Rural Microgrid Use in Emerging Markets*, 2017. [Online]. Available: <https://www.greentechmedia.com/articles/read/what-smart-meters-are-telling-us-about-microgrid-use?> (visited on 11/30/2018).
- [90] K. Louw, B. Conradie, M. Howells, and M. Dekenah, "Determinants of electricity demand for newly electrified low-income African households", *Energy Policy*, vol. 36, no. 8, pp. 2814–2820, 2008.
- [91] B. Otieno, N. J. Williams, and P. Mcsharry, "Customer segmentation for East African Microgrid Consumers", in *2018 IEEE PES/IAS PowerAfrica*, IEEE, 2018, pp. 468–473.
- [92] M. Zeyringer, S. Pachauri, E. Schmid, J. Schmidt, E. Worrell, and U. B. Morawetz, "Analyzing grid extension and stand-alone photovoltaic systems for the cost-effective electrification of Kenya", *Energy for Sustainable Development*, vol. 25, pp. 75–86, 2015.
- [93] D. Vercamer, B. Steurtewagen, D. Van Den Poel, and F. Vermeulen, "Predicting Consumer Load Profiles Using Commercial and Open Data", *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3693–3701, 2016.
- [94] J. L. Viegas, S. M. Vieira, R. Melício, V. M. Mendes, and J. M. Sousa, "Classification of new electricity customers based on surveys and smart metering data", *Energy*, vol. 107, pp. 804–817, 2016.
- [95] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, "Clustering analysis of residential electricity demand profiles", *Applied Energy*, vol. 135, pp. 461–471, 2014.
- [96] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping", *Energy*, vol. 42, no. 1, pp. 68–80, 2012.
- [97] J. P. Gouveia and J. Seixas, "Unraveling electricity consumption profiles in households through clusters: Combining smart meters and door-to-door surveys", *Energy and Buildings*, vol. 116, pp. 666–676, 2016.
- [98] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data", *Applied Energy*, vol. 141, pp. 190–199, 2015.
- [99] N. J. Williams and P. Jaramillo, "Electricity consumption and load profile segmentation analysis for rural microgrid customers in Tanzania", in *2018 IEEE PES/IAS PowerAfrica*, IEEE, 2018, pp. 360–365.
- [100] J. Llanos, R. Morales, A. Núñez, D. Sáez, M. Lacalle, L. G. Marín, R. Hernández, and F. Lanas, "Load estimation for microgrid planning based on a self-organizing map methodology", *Applied Soft Computing Journal*, vol. 53, pp. 323–335, 2017.

- [101] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, “Two-stage pattern recognition of load curves for classification of electricity customers”, *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1120–1128, 2007.
- [102] K. Mets, F. Depuydt, and C. Develder, “Two-Stage Load Pattern Clustering Using Fast Wavelet Transformation”, *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2250–2259, 2016.
- [103] M. Amin-Naseri and A. Soroush, “Combined use of unsupervised and supervised learning for daily peak load forecasting”, *Energy Conversion and Management*, vol. 49, no. 6, pp. 1302–1308, 2008.
- [104] J. Williams, “Clustering Household Electricity Use Profiles”, *Proceedings of Workshop on Machine Learning for Sensory Data Analysis*, pp. 19–26, 2013.
- [105] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, Applications*. Alexandria: SIAM ASA, 2007.
- [106] G. Chicco, R. Napoli, and F. Piglione, “Comparisons Among Clustering Techniques for Electricity Customer Classification”, *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 1–7, 2006.
- [107] L. Jin, D. Lee, A. Sim, S. Borgeson, K. Wu, C. A. Spurlock, and A. Todd, “Comparison of Clustering Techniques for Residential Energy Behavior Using Smart Meter Data”, in *AAAI Workshop on Artificial Intelligence for Smart Grids and Smart Buildings*, 2017, pp. 260–266.
- [108] K. Keitsch, H. Kondziella, and T. Bruckner, “Methodology for extracting dynamic standard load profiles from smart meter data”, in *Symposium Energieinnovation*, 2016, pp. 1–11.
- [109] W. Qiu, F. Zhai, Z. Bao, B. Li, Q. Yang, and Y. Cao, “Clustering approach and characteristic indices for load profiles of customers using data from AMI”, in *China International Conference on Electricity Distribution, CIGED*, 2016.
- [110] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, and M. Kolehmainen, “Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data”, *Applied Energy*, vol. 87, no. 11, pp. 3538–3545, 2010.
- [111] F. Rodrigues, J. Duarte, V. Figueiredo, Z. Vale, and M. Cordeiro, “A Comparative Analysis of Clustering Algorithms Applied to Load Profiling”, in *Machine Learning and Data Mining in Pattern Recognition*, 2003, pp. 73–85.
- [112] I. Benítez, A. Quijano, J. L. Díez, and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers”, *International Journal of Electrical Power and Energy Systems*, vol. 55, pp. 437–448, 2014.
- [113] D. Gerbec, S. Gašperič, I. Šmon, and F. Gubina, “Allocation of the load profiles to consumers using probabilistic neural networks”, *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 548–555, 2005.
- [114] K. L. Zhou, S. L. Yang, and C. Shen, “A review of electric load classification in smart grid environment”, *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103–110, 2013.
- [115] D. Arthur and S. Vassilvitskii, “K-Means++: the Advantages of Careful Seeding”, *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, vol. 8, pp. 1027–1025, 2007. arXiv: 1212.1121.
- [116] J. Kwac, J. Flora, and R. Rajagopal, “Household energy consumption segmentation using hourly data”, *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.
- [117] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On Spectral Clustering: Analysis and an Algorithm”, in *Advances in Neural Information Processing Systems*, vol. 14, Cambridge: MIT Press, 2002, pp. 849–856.

- [118] C. Alzate, M. Espinoza, B. De Moor, and J. A. K. Suykens, "Identifying customer profiles in power load time series using spectral clustering", in *Artificial Neural Networks – ICANN 2009. Lecture Notes in Computer Science*, C. Alippi, M. Polycarpou, C. Panayiotou, and G. Ellinas, Eds., vol. 5769, Berlin, Heidelberg: Springer, 2009, pp. 315–324.
- [119] Y. Weiss, "Segmentation using eigenvectors: a unifying view", *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 975–982, 1999.
- [120] B. J. Frey and D. Dueck, "Clustering by passing messages between data points", *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
- [121] V. K. Nguyen, W. E. Zhang, Q. Z. Sheng, and J. Merefield, "Mining Load Profile Patterns for Australian Electricity Consumers", in *Advanced Data Mining and Applications*, G. Cong, W.-C. Peng, W. E. Zhang, C. Li, and A. Sun, Eds., Cham: Springer International Publishing, 2017, pp. 781–793.
- [122] A. K. Zarabie, S. Lashkarbolooki, S. Das, K. Jhala, and A. Pahwa, "Load Profile Based Consumer Clustering Using Affinity Propagation", in *In Process*, 2018.
- [123] T. Kohonen, *Self-organization and associative memory*, 3rd Editio. Berlin, Heidelberg, Germany: Springer-Verlag, 1989.
- [124] G. Chicco, R. Napoli, F. Piglion, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers", *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1232–1239, 2004.
- [125] R. Granell, C. J. Axon, and D. C. Wallom, "Clustering disaggregated load profiles using a Dirichlet process mixture model", *Energy Conversion and Management*, vol. 92, pp. 507–516, 2015.
- [126] U. von Luxburg, R. C. Williamson, and I. Guyon, "Clustering: Science or Art?", in *JMLR: Workshop and Conference Proceedings*, vol. 27, 2012, pp. 65–79.
- [127] N. Yamaguchi, J. Han, G. Ghatikar, S. Kiliccote, M. A. Piette, and H. Asano, "Regression models for demand reduction based on cluster analysis of load profiles", *IEEE PES/IAS Conference on Sustainable Alternative Energy (SAE)*, pp. 1–7, 2009.
- [128] B. Barton, S. Blackwell, G. Carrington, R. Ford, R. Lawson, J. Stephenson, P. Thorsnes, and J. Williams, "Energy Cultures: Implication for Policymakers", University of Otago, Dunedin, Tech. Rep. February, 2013, p. 42.
- [129] S. Walton, A. Doering, C.-a. Gabriel, and R. Ford, "Energy Transitions: Lighting in Vanuatu (Project Report)", Tech. Rep., 2014.
- [130] D. Hopkins and J. Stephenson, "The replication and reduction of automobility: Findings from Aotearoa New Zealand", *Journal of Transport Geography*, vol. 56, pp. 92–101, 2016.
- [131] J. Stephenson, B. Barton, G. Carrington, D. Hopkins, M. Lavelle, R. Lawson, D. Rees, M. Scott, P. Thorsnes, S. Walton, and B. Wooliscroft, "Energy Cultures Policy Briefs (Project Report)", University of Otago: Centre for Sustainability, Tech. Rep., 2016.
- [132] S. Buttrey and L. Whitaker, *A Data Scientist's Guide to Acquiring, Cleaning, and Managing Data in R*. Chichester: John Wiley & Sons, Ltd, 2017.
- [133] K. Martinez, R. Ong, and J. Hart, "Glacsweb: a sensor network for hostile environments", *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, pp. 81–87, 2004.
- [134] P. Thomson, R. Hope, and T. Foster, "GSM-enabled remote monitoring of rural handpumps: a proof-of-concept study", *Journal of Hydroinformatics*, vol. 14, no. 4, p. 829, 2012.

- [135] H. Louie, G. Goldsmith, P. Dauenhauer, and R. H. Almeida, “Issues and applications of real-time data from off-grid electrical systems”, *IEEE PES PowerAfrica Conference*, pp. 88–92, 2016.
- [136] J. R. Vig, “Introduction to Quartz Frequency Standards”, Army Research Laboratory, Electronics and Power Sources Directorate, Fort Monmouth, NJ, USA, Tech. Rep., 1992.
- [137] Y. C. Wu, Q. Chaudhari, and E. Serpedin, “Clock synchronization of wireless sensor networks”, *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, Jan. 2011.
- [138] F. Ingelrest, G. Barrenetxea, G. Schaefer, M. Vetterli, O. Couach, and M. Parlange, “SensorScope: Application-Specific Sensor Network for Environmental Monitoring”, *ACM Transactions on Sensor Networks*, vol. 6, no. 2, pp. 1–32, 2010.
- [139] Freescale, *K64 Sub-Family Reference Manual Rev. 2*, 2014.
- [140] Datasheet, *ECX-31B SMD Tuning For Crystal Datasheet*.
- [141] E. McKenna, I. Richardson, and M. Thomson, “Smart meter data: Balancing consumer privacy concerns with legitimate applications”, *Energy Policy*, vol. 41, pp. 807–814, 2012.
- [142] A. I. Elombo, T. Morstyn, D. Apostolopoulou, and M. D. McCulloch, “Residential load variability and diversity at different sampling time and aggregation scales”, *2017 IEEE AFRICON: Science, Technology and Innovation for Africa*, pp. 1331–1336, 2017.
- [143] L. Sachs, *Applied Statistics. A Handbook of Techniques*. Berlin: Springer, 1984.
- [144] G. W. Irwin, W. Monteith, and W. C. Beattie, “Statistical electricity demand modelling from consumer billing data”, *IEE Proceedings C- Generation, Transmission and Distribution*, vol. 133, no. 6, p. 328, 1986.
- [145] International Energy Agency, “Africa Energy Outlook. A focus on energy prospects in Sub-Saharan Africa. World Energy Outlook Special Report”, Paris, Tech. Rep., 2014.
- [146] C. K. Lee, B. G. Bhang, D. K. Kim, S. H. Lee, H. L. Cha, and H. K. Ahn, “Estimation of Load Pattern for Optimal Planning of Stand-Alone Microgrid Networks”, *Energies*, vol. 11, no. 8, 2018.
- [147] M. Franz, N. Peterschmidt, M. Rohrer, and B. Kondev, *Mini-grid Policy Toolkit: Policy and Business Frameworks for Successful Mini-grid Roll-outs*, Eschborn, 2014.
- [148] M. Pipattanasomporn, M. Kuzlu, S. Rahman, and Y. Teklu, “Load profiles of selected major household appliances and their demand response opportunities”, *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 742–750, 2014. arXiv: 0603039 [quant-ph].
- [149] H. Olk and J. Mundt, “Photovoltaics for Productive Use Applications: A Catalogue of DC-Appliances”, GIZ, Tech. Rep., 2016, p. 252.
- [150] *Historical Average Weather*. [Online]. Available: <https://worldweatheronline.com> (visited on 01/01/2019).