

CRANFIELD UNIVERSITY

TOLULOPE OLUMUYIWA FALOPE

AN IMPROVED ENERGY MANAGEMENT SYSTEM FRAMEWORK
FOR SOLAR ENERGY INTEGRATION

SCHOOL OF WATER, ENERGY AND ENVIRONMENT
PhD in Energy and Sustainability

PhD
Academic Year: 2021 - 2024

Supervisor: Dr. Liyun Lao
Associate Supervisor: Dr. Da Huo
May 2024

CRANFIELD UNIVERSITY

SCHOOL OF WATER, ENERGY AND ENVIRONMENT
PhD in Energy and Sustainability

PhD

Academic Year 2021 - 2024

TOLULOPE OLUMUYIWA FALOPE

AN IMPROVED ENERGY MANAGEMENT SYSTEM FRAMEWORK
FOR SOLAR ENERGY INTEGRATION

Supervisor: Dr. Liyun Lao
Associate Supervisor: Dr. Da Huo
May 2024

This thesis is submitted in partial fulfilment of the requirements for
the degree of PhD

© Cranfield University 2024. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

ABSTRACT

Renewable energy sources like wind and solar play a crucial role in decarbonizing energy supply, but their variable and intermittent nature lead to reliability and stability issues. One way of sustainably integrating these energy sources into the grid is through an energy management system. The study reported in this thesis gives a comprehensive definition of an integrated energy management system and creates a novel framework that identifies energy forecasting, demand-side management, and supply-side management, as crucial components for grid balancing. In addition, this research looks particularly at solar integration, and how the integrated energy management system offers a unique combination of solar energy forecasting, time-of-use tariffs, direct load control demand response, and generator control, in increasing penetration levels of solar energy. The significance of this research is that the proposed system presents a viable, sustainable, and cheaper way of increasing PV usage and thereby grid penetration by prioritising efficient use of available PV supply before calling up additional supply. To validate the proposed integrated energy management system, this research looks to understand the functions of each individual component and how their interconnectedness creates a novel management system.

Firstly, this research develops a three-step solar forecasting approach that uses low-level data fusion to combine weather variables from both an on-site and a local weather station to improve solar energy forecasting. The forecasting model response is historic PV generation, and the predictors are weather variables with moderate to strong positive correlations to solar radiation. Data obtained is preprocessed using Low-level Data Fusion, Pearson Correlation Coefficient analysis, Rescaling method, and List-wise Deletion method. This approach is then tested on a 1MW utility scale solar plant, resulting in a 6% and 13% prediction accuracy improvement when compared to solely using data from an on-site, and local weather stations respectively. This approach is also validated for three residential rooftop solar systems (8 kW, 10.5 kW and 15 kW), achieving root mean square error values of 0.0984, 0.1425, and 0.0885

respectively. The resulting low root mean square error values, a measure of the predicted PV to actual PV generation, proves that the model can be adopted for different PV plant sizes and is suitable for any customer across the distributed generation spectrum. To further improve the accuracy of the model, other preprocessing techniques are investigated and applied. The study shows that the combination of Low-level Data Fusion, Linear Interpolation, filling outliers, data smoothing, Rescaling method, moderate to strong PV correlation of weather parameters using Pearson Correlation Coefficient, day/time/month decomposition, seasonal decomposition, Principal Component Analysis, and holdout validation, increases the accuracy of the model by 75%.

The ability of direct load control to manage energy consumption is validated in a case study by using Connected Power's unique smart sockets and Lumen radio's Mira Mesh Radio Frequency wireless network. Small plug-in loads were connected to ten smart sockets located in a robotics laboratory and a café, resulting in reduced energy consumption by 44% and 72% respectively when compared to the baseline without direct load control.

Finally, the integrated energy management system framework is validated by testing its capacity to increase PV usage for an off-grid residential house with a PV/diesel generator power source. A decision-based algorithm is created that adjusts PV supply forecast errors, initiates direct load control responses to reduce excess load during periods of low PV supply, and/or increase power supply by calling up a diesel generator. In addition, this is combined with the proposed three-step solar energy forecasting approach and a programmable load schedule based on time-of-use criteria. The effects of customer behaviour are also analysed by using a 14% override rate, with 80% preconditioning and 20% rebounding. The hybrid PV/diesel generator power source with the proposed integrated energy management system is compared against two configurations: a baseline configuration that uses a solely diesel generator source, and a hybrid PV/diesel generator power source. Results show that the integrated energy management system reduced the lifetime expenditure costs and CO₂ emissions by 44% and 46% respectively when compared to the baseline configuration, and

by 8% and 9% in the hybrid photovoltaic/diesel generator, while also increasing the PV usage from this configuration by over 113%.

This research also addresses opportunities and limitations of the proposed system and lays the foundation for future research using other intermittent renewable energy sources such as wind.

Keywords:

Integrated energy management system; solar energy forecasting; demand side management; supply side management; time-of-use; direct load control

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to those who have contributed to the completion of this thesis.

First and foremost, I would like to thank the good Lord who supported me, provided for me, and gave me the strength to finish this thesis. I wouldn't have been able to do it without him.

Heartfelt appreciation goes to my wife, Ibukunoluwa, and son Jaden. Their love, devotion and sacrifice spurred me on. Special thanks to my parents, Dr. and Mrs. Falope for their financial support, encouragement, and prayers. To my in-laws, Dr. and Mrs. Ilesanmi and the rest of my family for their unwavering support, understanding, and encouragement throughout this academic journey.

I extend my sincere appreciation to my thesis supervisors. To Dr. Liyun Lao, Dr. Da Huo, and Dr. Dawid Hanak for their invaluable guidance, unwavering support, and insightful feedback throughout the research process. Their expertise and encouragement played a pivotal role in shaping the direction of this work.

I would like to express my sincere gratitude to the Petroleum Technology Development Fund for their generous financial support that made this research possible. Their commitment to advancing academic studies and promoting research excellence has been instrumental in the successful completion of this project.

I would also like to acknowledge the School of Water, Energy and Environment for providing a conducive environment for academic growth and research. The resources and facilities offered were instrumental in the successful completion of this thesis.

Finally, I want to express my gratitude to my friends for their moral support and occasional distractions that provided much-needed breaks during intense research periods.

This thesis would not have been possible without the collective efforts of these individuals. Each one has played a crucial role in shaping this work, and I am truly grateful for their contributions.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS.....	v
LIST OF FIGURES.....	xi
LIST OF TABLES	xiii
LIST OF ABBREVIATIONS.....	xv
1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Motivation	3
1.3 Aims and objectives	7
1.4 Significance of this research	7
1.5 Interconnection of chapters and project outputs	8
1.6 Dissemination from the PhD thesis	1
1.6.1 Peer-reviewed journal publications	1
1.6.2 Conference publications.....	1
References	2
2 Hybrid energy system integration and management for solar energy: A review	4
2.1 Introduction	5
2.2 Methods	9
2.2.1 Solar energy forecasting	10
2.2.2 Energy management system.....	11
2.2.3 Supply side management.....	11
2.2.4 Demand Side Management.....	12
2.3 Renewable smart hybrid mini-grids: The future of renewable energy integration.....	13
2.3.1 Distributed generation: Mini-grid & renewable energy home systems.....	14
2.3.2 Renewable mini-grids and hybrids	18
2.3.3 Grid-tied versus off-grid.....	18
2.3.4 Renewable smart hybrid mini-grid.....	19
2.4 Solar energy forecasting	20
2.5 Demand side management.....	28
2.5.1 Demand response.....	30
2.5.2 Price-based demand response program	32
2.5.3 Incentive-based demand response program	33
2.6 Supply side management	35
2.6.1 Back-up power and supply side management: Batteries and generator dispatch	36
2.6.2 Flexible power supply generation.....	38
2.7 Energy management system	38

2.7.1 Energy management system control	39
2.7.2 Smart meters	41
2.8 Integrated energy management system	41
2.9 IEMS perspective and future research	54
2.10 Conclusion	58
References	60
3 A Three-step weather data approach in solar energy prediction using machine learning	92
3.1 Introduction	93
3.2 Data source and type: A Literature review	96
3.3 Methods	99
3.3.1 Data description	100
3.3.2 Data comparison	101
3.3.3 Data pre-processing	102
3.3.4 Model performance evaluation	102
3.3.5 Regression learner app methodology and flow chart	103
3.4 Methodology	104
3.4.1 Forecasting architecture	105
3.4.2 Case study description	106
3.5 Results and discussion	106
3.5.1 Cranfield correlation matrix	107
3.5.2 Bedford correlation matrix	110
3.5.3 Performance of forecasting architecture	111
3.5.4 Validation of forecasting architecture	118
3.6 Conclusion	119
References	120
4 Data preprocessing techniques for statistical PV forecasting models	133
4.1 Introduction	134
4.2 Data preprocessing methods	136
4.2.1 Data cleaning	137
4.2.2 Data reduction	139
4.2.3 Data decomposition	142
4.2.4 Data scaling	143
4.2.5 Data integration	144
4.2.6 Data partitioning	144
4.3 Three-step forecasting model	145
4.4 Model evaluation	145
4.4.1 Validation methods	145
4.4.2 Performance metrics	146
4.5 Case study	148
4.6 Results and discussion	149
4.7 Conclusion	163

References	165
5 Direct Load Control approach: An IoT-based plug load management system for smart buildings	176
5.1 Introduction	177
5.2 Literature review	179
5.3 Methods	181
5.3.1 System description	181
5.3.2 Baseline comparison method	183
5.3.3 CO ₂ emissions	185
5.3.4 Payback period	185
5.3.5 Case study	185
5.4 Results and discussion	187
Conclusion	195
References	196
6 Integrated energy management system for off-grid solar energy application	202
6.1 Introduction	203
6.2 Methodology	211
6.2.1 Three-step forecasting architecture	212
6.2.2 Financial forecasting tools	213
6.2.3 Dispatch strategies	217
6.2.4 Integrated energy management system rule-based algorithm	218
6.3 Case study	221
6.3.1 Data description	221
6.3.2 Financial and economic analysis	222
6.4 Results and discussion	227
6.4.1 Percent share of diesel generator and photovoltaic loads	227
6.4.2 Forecast accuracy and algorithm error correction	229
6.4.3 Lifetime expenditure and CO ₂ emissions	229
6.4.4 Techno-economic analysis	232
Conclusion	234
References	236
7 Discussions, conclusions, and recommendations	246
7.1 Discussions	246
7.2 Conclusion	254
7.2.1 Three-step forecasting model development	254
7.2.2 Direct load control: A case study	256
7.2.3 Deterministic analysis of the integrated energy management framework	256
7.2.4 Key contributions to knowledge	257
7.2.5 Research limitations	258
7.3 Recommendations for future research	259

References	261
APPENDICES	266
Appendix A A Three-step weather data approach in solar energy prediction using machine learning	266
Appendix B Data preprocessing techniques for statistical PV forecasting models	269

LIST OF FIGURES

Figure 1-1 Share of cumulative power capacity by technology form 2010 – 2027 [3]	2
Figure 1-2 Global installed capacity and electricity generation of renewable energy sources from 2013 – 2022	2
Figure 1-3 Structure of the PhD thesis and linkage of chapters and manuscripts to the thesis objectives	1
Figure 2-1 Review approach for IEMS Framework.....	13
Figure 2-2 Power system configurations for energy management system deployment.....	14
Figure 2-3 Classification of distributed generation systems	17
Figure 2-4 Demand side management architecture	29
Figure 2-5. Demand response architecture	31
Figure 3-1 Flow chart of the regression learner app methodology	104
Figure 3-2 Proposed three-step forecasting architecture	106
Figure 3-3 2019 Weather parameters from the Kisanhum Weather Station in Cranfield University [62].....	109
Figure 3-4 Case study: Three-step forecasting architecture block diagram ...	112
Figure 3-5 Model performance of Boosted Trees ((a) Scenario 1 (b) Scenario 2 (c) Scenario 3).....	115
Figure 3-6 Actual and predicted PV generation on 12th September 2019 ((a) Scenario 1 (b) Scenario 2 (c) Scenario 3).....	116
Figure 3-7 Actual and predicted PV generation for Scenario 3 on the 37th week in 2019.....	118
Figure 4-1 Data preprocessing framework for statisitcal PV forecasting models	137
Figure 4-2 Three-Step Forecasting Architecture	145
Figure 4-3 Modified flow chart of the Regression learner app	149
Figure 4-4 Weather data from Kisanhub, Cranfield University showing the inverse relationship between Solar Radiation and Relative Humidity in 2019.....	156
Figure 4-5 Model Performance of ((a) Solar Radiation correlation (b) Photovoltaic correlation).....	157
Figure 4-6 Flow chart summary of data preprocessing iterations.....	160

Figure 4-7 One-week prediction comparison between original and improved Three-step forecasting method.....	162
Figure 5-1 Network architecture for Connected Power.....	182
Figure 5-2 Baseline comparison method for estimating energy savings	184
Figure 5-3 Small power plug-in loads connected to MK Electric smart sockets	186
Figure 5-4 Robotics laboratory half-hourly load consumption - 01/08/2023 ...	187
Figure 5-5 Device consumption comparison in kWh	192
Table 6-16-1 Comparison of Studies on Various EMS Frameworks	209
Figure 6-6-2 Proposed Integrated Energy Management System Framework	211
Figure 6-3 Block diagram of proposed IEMS.....	212
Figure 6-4 Proposed Three-Step Forecasting Approach.....	213
Figure 6-5 Flowchart for proposed IEMS.....	218
Figure 6-6 One-day load profile of residential house.....	222
Figure 6-7 Demand Load Control/Time-of-use Program	226
Figure 6-8 Comparison of DG/PV and DG/PV/IEMS configurations servicing the load.....	228
Figure 6-9 Cost comparison of DG, DG/PV, and DG/PV/IEMS configurations	232
Figure 7-1 Proposed solar-based IEMS framework.....	251

LIST OF TABLES

Table 1-1 Chapter and paper organisation	Error! Bookmark not defined.
Table 2-1 Recently published reviews related to solar energy management system and integration.....	8
Table 2-2 Review of research studies on data source and types	21
Table 2-3 A summary of studies on solar energy forecasting.....	23
Table 2-4 Comparison of demand response programs	34
Table 2-5 A summary of studies on integrated energy management systems .	48
Table 3-1 Summary of day-ahead solar energy forecasting studies.....	95
Table 3-2 PV system specifications of 1 MW solar farm in Cranfield University	100
Table 3-3 PV system specification of three roof-top solar arrays in Cranfield University.....	101
Table 3-4 Correlation correlation matrix	110
Table 3-5 Correlation matrix of weather parameters in Bedford	110
Table 3-6 RLA model training input parameters	112
Table 3-7 Performance comparison of regression models in Regression learner app for Scenario 3	112
Table 3-8 Performance comparison of case study scenarios using Boosted Trees	114
Table 3-9 Validation results of proposed architecture using residential solar arrays.....	118
Table 4-1 Study comparisons on data preprocessing methods.....	135
Table 4-2 Performance of data preprocessing methods.....	150
Table 4-3 Correlation matrix of weather Parameters in Cranfield and Bedford weather stations to ((a) Solar radiation (b) Photovoltaic generation)	154
Table 4-4 Performance of Solar Radiation correlation and Photovoltaic correlation	156
Table 4-5 Comparing performance of original Three-step to improved Three-step forecasting model	161
Table 5-1 Summary of direct load control studies and end-user load.....	180
Table 5-2. Studies showing direct load control architecture	182

Table 5-3 Socket and outlet designations	185
Table 5-4 Results of the Baseline comparison method	189
Table 5-5 Results of techno-economic analysis	194
Table 6-16-1 Comparison of Studies on Various EMS Frameworks	209
Table 6-2 Financial conditions and definitions.....	216
Table 6-3 Load demand summary of residential house.....	222
Table 6-4 Parameter assumptions	226
Table 6-5 Consumption comparison between DG, DG/PV, and DG/PV/IEMS configurations for a residential house in 2019	227
Table 6-6 Fuel comparison between DG, DG/PV, and DG/PV/IEMS configurations	230
Table 6-7 Cost, Output and CO2 emissions comparison between DG, DG/PV, and DG/PV/IEMS configurations	231
Table 6-8 Techno-economic analysis of DG/PG and DG/PV/IEMS configurations	233
Table 7-1 Summary of novelty and findings	252

LIST OF ABBREVIATIONS

AC	Air Conditioners
AMI	Advanced Metering Infrastructure
ANN	Artificial Neural Network
ASD	Atomic Sparse Decomposition
BACnet/IP	Building Automation and Control Network over Internet Protocol
BEMS	Building Energy Management System
BESS	Battery Energy Storage System
BiLSTM	Bidirectional Long Short-Term Memory
CCHP	Combined Cooling, Heat and Power
CEMS	Centralized Energy Management System
CoE	Cost of Energy
CP	Connected Power
CPP	Critical Peak Pricing
CRP	Cost-reflective Pricing
DF	Daylight-free
DFT	Discrete Fourier Transform
DDG	Decentralized Distributed Generation
DG	Diesel Generator
DLC	Direct Load Control
DO	Delete Outliers
DP	Data Preprocessing
DR	Demand Response
DS	Data Smoothing
DSM	Decimal Score Method
DSM	Demand Side Management
DSP	Distribution System Planning
ED	Energy Dispatch
EMD	Empirical Mode Decomposition
EMS	Energy management system
ENN	Elman Neural Network
ESS	Energy storage system
EV	Electric Vehicle

EWT	Empirical Wavelet Transform
FO	Fill Outliers
FRNN	Full Recurrent Neural Network
FWPD	Full Wavelet Packet Decomposition
GHG	Greenhouse gas
GPR	Gaussian Process Regression
GRU	Gated Recurrent Unit Network
HEMS	Home Energy Management System
HVAC	Heating, Ventilation, and Air Conditioner
IDR	Integrated Demand Response
IEA	International Energy Agency
IEMS	Integrated Energy Management System
IES	Integrated Energy System
IoT	Internet-of-things
IRENA	International Renewable Agency
ISEMS	Intelligent Smart Energy Management System
KPCA	Kernel Principal Component Analysis
LCoE	Levelized Cost of Energy
LIM	Linear Interpolation Method
LSTM	Long Short-Term Memory
LWDM	List-wise Deletion Method
MAE	Mean Absolute Error
MG	Mini-Grid
MG SVM	Medium Gaussian Support Vector Machine
MM	Moving Mean
MQTT	Message Queueing Telemetry Transport
MSE	Mean Squared Error
NC	Negative Correlation
NERC	Nigerian Electricity Regulatory Commission
NMSE	Normalized Mean Squared Error
NN	Neural Networks
NPV	Net Present Value
NWP	Numerical Weather Predictions

PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PEMS	Predictive Energy Management System
PV	Photovoltaic
PVC	Photovoltaic Correlation
REHS	Renewable Energy Home System
REMS	Real-time Energy Management System
RES	Renewable Energy Sources
RHMG	Renewable Hybrid Mini-grid
RLA	Regression Learner App
RM	Rescaling Method
RMG	Renewable Mini-grid
RMSE	Root Mean Square Error
RRVFLN	Robust Minimum Variance Random Vector Functional Link Network
RSHMG	Renewable Smart Hybrid Mini-grid
RTP	Real-time Pricing
SEF	Solar Energy Forecasting
SEMS	Smart Energy Management System
SHS	Solar Home System
SM	Smart Meters
SoC	State of Charge
SR	Solar Radiation
SR	Supply Response
SRC	Solar Radiation Correlation
SSM	Supply Side Management
SVM	Support Vector Machine
TCLs	Thermostatically Controlled Loads
TOU	Time-of-Use
UC	Unit Commitment
VMD	Variational Mode Decomposition
VOCACs	Variable-Output Compressor Air Conditioners
VRE	Variable Renewable Energy
WD	Wavelet Decomposition

WNN	Wide Neural Network
XGBoost	Extreme Gradient Boosting
ZS	Z-score Method

1 INTRODUCTION

1.1 Background

According to British Petroleum's (BP) energy outlook 2023, global energy consumption will reach 667 EJ (exajoules) by 2025 and 760 EJ by 2050 [1]. Globally, energy is thought to be responsible for over three-quarters of all greenhouse gas (GHG) emissions according to the International Energy Agency (IEA) [2]. Fossil fuels accounted for 80% of the world's energy supply in 2021; oil made up about 30%, followed by coal (27%), natural gas (24%), and other fuels. Coal also accounted for 44% of global fuel combustion emissions, with oil coming in second at 32% and natural gas at 22% [2]. To mitigate the effects of climate change, the global energy landscape is undergoing a transformative shift from conventional fossil fuel-based power generation to renewable alternatives that have little to no carbon footprints. This trend is supported by the previously cited BP report, showing renewable energy consumption and conventional fossil fuels with a 61% increase and a 15% decrease between 2025 and 2050 respectively. Among the various renewable energy sources (RES), solar energy/Photovoltaic (PV) has emerged as a particularly promising option, offering a clean and abundant resource for electricity production. In another report by IEA [3], installed power capacity from solar PV is expected to surpass that of coal by 2027, having the highest capacity of any power source in the world as can be seen in Figure 1-1.

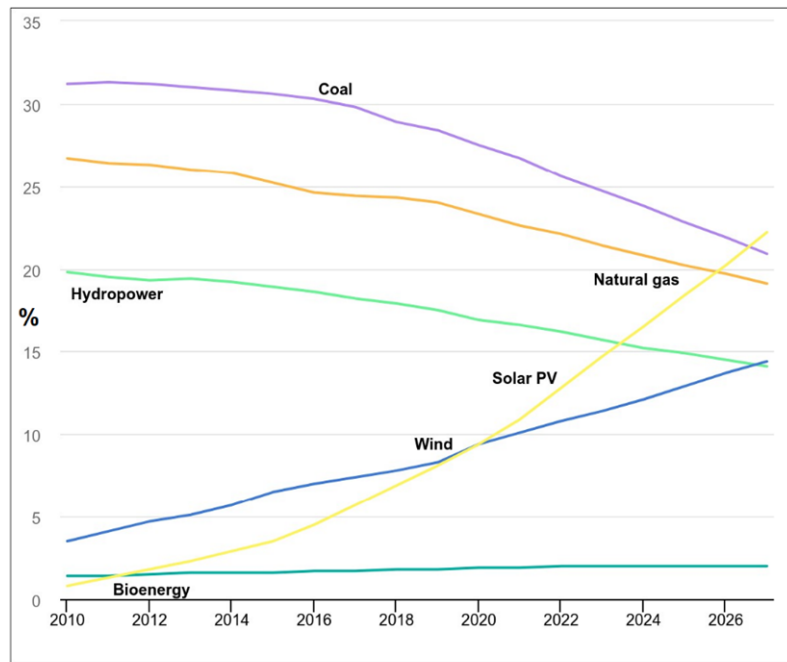


Figure 1-1 Share of cumulative power capacity by technology form 2010 – 2027 [3]

When looking at the maximum net electricity generating capacity of RES-installed power plants, a report by the International Renewable Energy Agency (IRENA) [4], shows that solar energy has seen the most increase in capacity and generation over the past decade of all the RES.

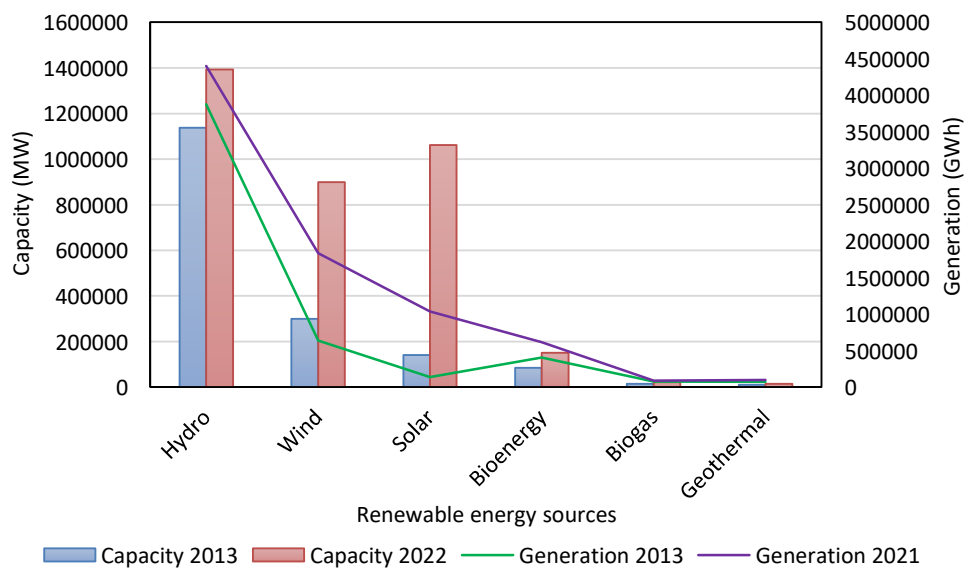


Figure 1-2 Global installed capacity and electricity generation of renewable energy sources from 2013 – 2022

According to Figure 1-2, solar energy has the highest increase in installed capacity (87%), followed by wind (67%) and bioenergy (44%). In terms of electricity generation, the trend is similar as solar energy comes in first (87%) followed by wind (65%) and bioenergy (34%). RES penetration keeps on increasing with solar, wind, and hydro contributing the largest share.

However, the variable and intermittent nature of solar energy creates challenges to grid integration causing reliability and stability problems [5]. This, along with today's complex multi-energy, multi-user, and multi-control grid, necessitates innovative approaches to harness the full potential of solar energy. The pertinent question becomes how can we efficiently and reliably increase solar energy penetration in today's grid? More specifically, the following knowledge gaps are to be addressed:

- Definition of an integrated energy management system (IEMS) as it relates to today's grid. There are a few existing definitions of IEMS, but none are expressed in terms of energy forecasting combined with the simultaneous control of demand and supply, and how indispensable these elements are in balancing today's grid.
- While Low-level Data Fusion (LLDF) is a well-researched approach, its application in enriching the data input source of a forecasting model, by combining weather data from the PV plant's location and another weather station in close proximity, has not been investigated.
- There are several IEMS configurations used in optimizing and increasing solar energy penetration. None however combine PV forecasting, time-of-use (TOU), direct load control (DLC), and generator control (GC) to match consumption with available power supply.

1.2 Motivation

Historically, the traditional grid had been largely characterised by homogeneous non-dispatchable power supply made up of conventional fossil fuels. As the grid expanded due to technological advancements, increased and varying customer profiles impacted how power would be generated, transmitted, and distributed.

The term “energy management system” (EMS) really began to trend after the energy crisis of 1973 and 1979 and is defined by the United Nations Industrial Development Organization (UNIDO) as a framework for energy consumers to manage their energy use [6]. EMS is a general term for the computer-aided instrumentation system used by electric utility grids and microgrid operators to monitor, regulate, and optimise the performance of the power system [7]. While the definition of EMS was derived in a time when the grid was largely supplied from “one-type source” with little or no contribution from RES, that definition would need to be broadened to fit today’s grid or a future one. Most utilities rely on EMS that were built in the early 1990s and are outdated, remaining largely unchanged and unaffected by advances in IT [8]. Today’s grid has evolved into a smart grid due to the introduction of the Internet of Things (IoT), big data, cloud computing, advanced metering infrastructure (AMI), and other communication and technological advancements [9,10].

The emergence of the modern grid gave birth to a new concept of integrated energy systems (IES). IES combines multiple energy sources like RES and fossil fuels by using advanced energy conversion technologies to efficiently convert them into usable forms of energy like heat and electricity [11]. IES enhance energy utilization while reducing carbon emissions [12], coordinating the planning and operational optimisation of heterogeneous systems [13]. Considering the EMS state-of-the-art and IES literature, integration is often defined as an optimization problem with set objective functions and constraints. The aim is to achieve an objective(s) like minimising costs or maximising profits, by developing an optimised process (for example efficient dispatch of power sources), while operating within constraints (for example installed capacity). In a typical example, Gao et al. [14] use an Analytic Hierarchy Process and the Fuzzy Comprehensive Evaluation method to optimally configure and select multi-energy storage systems based on power responses. Guo and Xiang [15] propose an interactive Integrated Energy System Planning (IESP) platform that aggregates various energy sources (electricity, gas, heat) and takes as its input investment decisions, load profiles, meteorological data, and techno-economic parameters converting into a mixed integer linear programming problem. The objectives were

to select the optimal component size of each energy source, optimise the total annual cost, and reduce carbon emissions. Feng et al. [16] considered the optimization of the IES under RES and energy demand uncertainties using the information gap decision theory method. He et al. [17] solve the annual cost of planning for an IES as the objective of its optimization problem by considering electric vehicles swapping stations and carbon capture power systems while using natural gas, heating networks, PV, and wind energy on the supply side. It is clear to see that the focus of these models is on energy conversion within constraints.

While EMS might be a general definition referring to both heterogeneous and homogeneous power sources, IES views energy integration solely as an optimization problem. In today's multi-user, multi-source and multi-control grid, there is a need to consider all aspects of integration including both the demand and supply sides. Since multi-source energy could also include intermittent sources like wind and solar energy, it is also necessary to consider energy forecasting. Therefore, a comprehensive framework for solar energy integration is proposed that combines solar energy forecasting (SEF), demand side management (DSM) and supply side management (SSM) to efficiently balance the load and power supply in the grid. This framework eliminates the deficiencies of an EMS and IES, combining both systems into an integrated energy management system (IEMS). This research endeavours to explore the synergistic relationship between solar energy integration and IEMS, examining their collective impact on creating a sustainable, reliable, and intelligent power infrastructure.

While IEMS is not a new concept, it has not been defined in terms of SEF, DSM and SSM. This research proposes the definition of an IEMS as a system that leverages advancements in technology and communication by integrating predictive and real-time controls to initiate both supply and demand responses in balancing the load and power supply in the grid. Like the name implies, SSM refers to the techniques used to control power supply to the grid. Examples of SSM are integrating battery storage, optimizing, and synchronizing alternative or

back-up power, ramping up and down of generation through automatic generator control (GC) and so forth. Consequently, DSM are techniques used to control and efficiently manage load demand. They include demand response (DR), energy operations and load scheduling, distributed generation and integrating energy storage. DR can be further broken down into price-based and incentive-based responses. Examples of price-based incentives include TOU pricing, critical peak pricing, and real-time pricing. Incentive based demand response includes DLC, demand bidding, and interruptible programs. The selection of GC as SSM, TOU and DLC as DSM is extensively discussed in Chapter 2.

Emerging from the novel definition of an IEMS in Chapter 2, this study seeks to validate the IEMS framework which leads to the following research questions:

- Will the solar forecasting model be physical, statistical or hybrid?
- What preprocessing methods will be used to ensure accurate input data into the model?
- Will the IEMS express solar intermittency as an optimization model, and if it does, will this be deterministic or stochastic, discrete or continuous?
- What is the contribution of each component in the IEMS and how are they interconnected?

This study will establish a clear definition and an understanding of today's IEMS and each of its components, with support and validation coming from case studies and literature references. The overall aim is to determine the IEMS's impact on the grid, particularly its ability to increase PV usage and grid penetration.

This study aspires to contribute to the growing body of knowledge on sustainable energy solutions by investigating the intricate relationship between solar energy integration and IEMS. By exploring theoretical frameworks, practical applications, and potential synergies, this research seeks to inform the ongoing discourse on creating resilient and intelligent power infrastructures that leverage the full potential of solar energy, paving the way for a more sustainable energy future.

1.3 Aims and objectives

The main aim of this research is to create and validate an IEMS that can maintain the demand-supply energy balance in solar-based power systems while increasing grid reliability and PV penetration to counter the effects of solar energy's intermittency. To achieve this aim, the following research objectives have been identified:

1. To assess different IEMS configurations and determine how individual components increase the reliability of solar-based power systems. It would lead to the identification of key elements and best practices for optimizing renewable energy systems, along with identifying potential knowledge gaps relating to methods and structure of solar-based IEMS.
2. To develop a method that uses LLDF in combining weather parameters from an on-site weather station and a local weather station to improve solar energy forecasting. This objective contributes to advancing the field of solar energy forecasting by exploring the potential benefits of LLDF. Such findings could be vital for optimizing solar power generation and reducing dependency on non-renewable energy sources.
3. To ascertain the efficacy of DLC in reducing energy wastage and consumption, thereby affirming its role as a feasible demand response strategy for optimizing the load to align with available supply.
4. Finally, to develop a comprehensive descriptor for assessing the impact of the proposed IEMS on solar hybrid applications, encompassing technical, economic, and environmental dimensions. This endeavour is to deepen insights into the holistic implications of the IEMS, spanning technical feasibility, economic viability, and environmental sustainability.

1.4 Significance of this research

This research creates and validates a holistic framework for solar energy integration by identifying key elements of the solar-based IEMS which are SEF, SSM and DSM. The effects of this could have a lasting impact on the environment, reducing CO₂ and other harmful GHG emissions. In addition, the

IEMS potentially creates a “virtual energy” source created by efficiently managing load due to DSM. The implication of this is that operators can build smaller mini grid systems and eliminate/reduce energy waste resulting in great cost savings for both on-grid and off-grid applications. It also presents a viable, sustainable, and cheaper way of extending the grid and increasing energy access to hard-to-reach places. This research also advances the field of SEF by developing a novel three-step forecasting architecture that uses LLDF to combine weather parameters from a reliable local weather station and an on-site weather station and identifies weather variables with a moderate to strong positive correlation to solar radiation to predict the next-day ahead PV generation. This is vital in optimizing solar energy and reducing reliability issues due to intermittency.

1.5 Interconnection of chapters and project outputs

Four themes link the thesis objectives, chapters, and publications as shown in Figure 1-3. These are literature review, case study analysis, model development and deterministic analysis. The thesis background and motivation were provided in Chapter 1. Chapter 2 (publication 1) provides a general literature review of the current state of solar energy integration, highlighting the theoretical foundations of IEMS and elucidating its core components and principles, with a specific focus on how the interconnectedness of SEF, SSM and DSM can facilitate the seamless integration of solar energy. The answer to this research question informs the creation of a novel IEMS framework that combines all the components mentioned above. Chapters 3 – 6, focuses on model development and performance validation of the IEMS. To address the research gap of a novel IEMS, the literature review begins by identifying the importance of forecasting models in VRE integration. This is addressed in Chapter 3 (manuscript 1) and Chapter 4 (manuscript 2), where the proposed SEF model is developed, improved, tested, and validated for both utility-scale and residential solar energy systems. In Chapter 5 (manuscript 3), a case study analysis is performed to test the effectiveness of DLC technology in the energy management of some small power plug-in loads at Cranfield University. The results from this chapter are important in validating DLC as a DSM response which is one of the foundations

of the IEMS highlighted in Chapter 2. The outputs of Chapters 2 – 5 contribute to the development of the IEMS framework in Chapter 6 (publication 2), where the SEF model developed in chapter 3 serves as the input into the predictive control of the IEMS while the DLC technique of Chapter 5 is validated as part of the real-time control. In addition, the findings of the literature review in Chapter 2 inform the selection of TOU and DLC as examples of DSM, while GC is selected as part of the SSM. A deterministic analysis is also carried to validate the IEMS framework in a solar hybrid configuration. Chapter 7 synthesizes all the thesis outputs and highlights the contribution to knowledge, draws up important conclusions, and makes recommendations for future work.

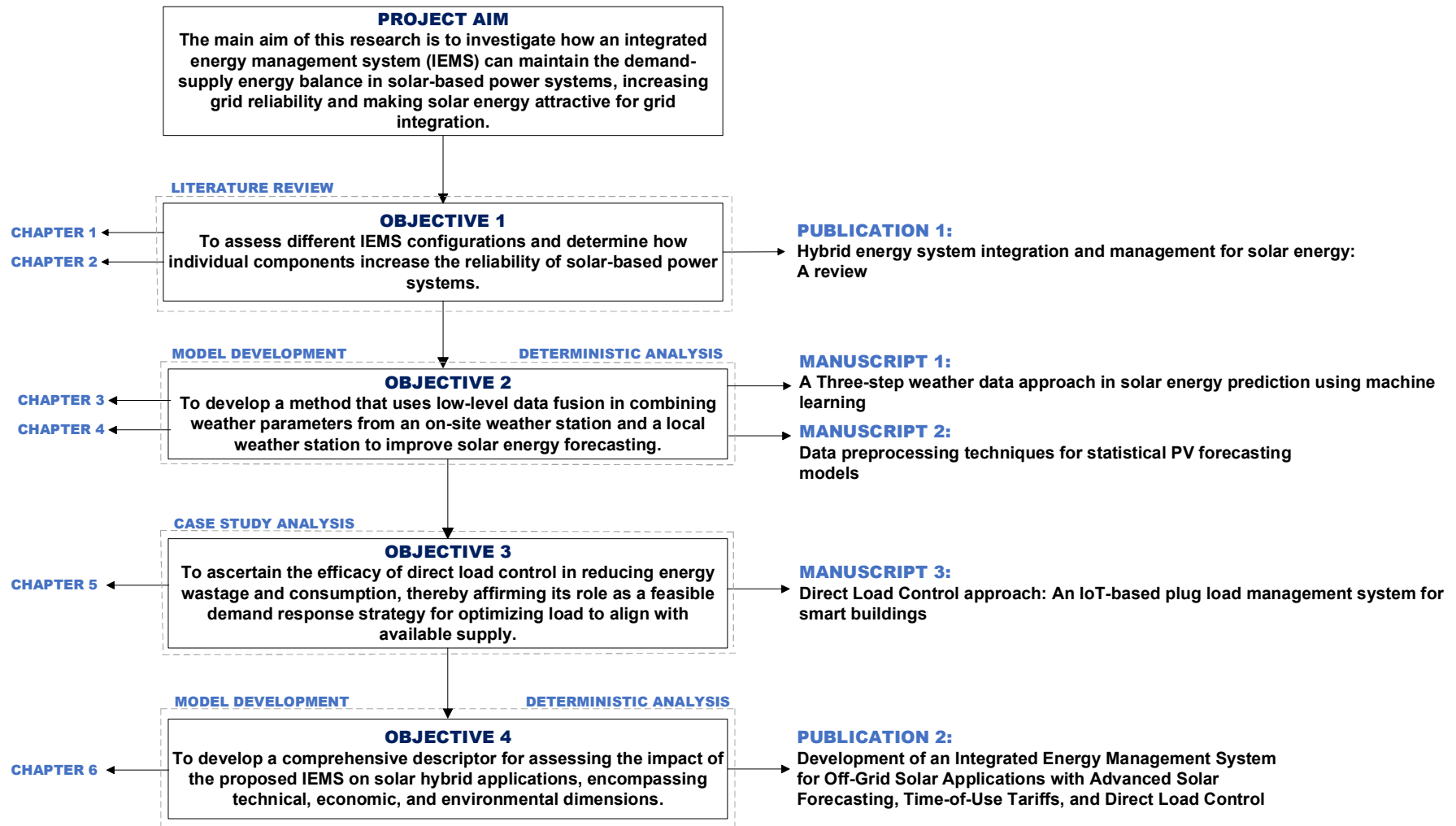


Figure 1-3 Structure of the PhD thesis and linkage of chapters and manuscripts to the thesis objective

1.6 Dissemination from the PhD thesis

1.6.1 Peer-reviewed journal publications

1. Falope T, Lao L, Hanak D, Huo D. Hybrid energy system integration and management for solar energy: A review. Energy Conversion and Management: X 2024;21:100527. <https://doi.org/10.1016/J.ECMX.2024.100527>.
2. Falope TO, Lao L, Huo D, Kuang B. Development of an integrated energy management system for off-grid solar applications with advanced solar forecasting, time-of-use tariffs, and direct load control. Sustainable Energy, Grids and Networks 2024;39:101449. <https://doi.org/10.1016/J.SEGAN.2024.101449>.

1.6.2 Conference publications

1. Falope TO, Lao L, Hanak D. A Unique Three-Step Weather Data Approach in Solar Energy Prediction Using Machine Learning. Sustainable Energy Solutions for a Post-COVID Recovery towards a Better Future: Part VII, Bangkok: International Conference on Applied Energy; 2021, p. 1–7. <https://doi.org/https://doi.org/10.46855/energy-proceedings-9863>.
2. Falope, T. O., Lao, L., & Huo, D, (2024) Integrated Energy Management System Approach for Off-Grid Residential Home. In: Future Directions in Energy Engineering. Springer, Cham. <https://doi.org/10.1007/978-3-031-62041-6>

References

- [1] BP. Energy Outlook: 2023 Edition. Bp Energy Economics 2023:1–66. <https://www.bp.com/en/global/corporate/energy-economics/energy-outlook.html> (accessed April 10, 2024).
- [2] IEA. Greenhouse Gas Emissions from Energy Data Explorer. Paris: 2023.
- [3] IEA. Share of cumulative power capacity by technology, 2010-2027. IEA 2022. <https://www.iea.org/data-and-statistics/charts/share-of-cumulative-power-capacity-by-technology-2010-2027> (accessed December 4, 2023).
- [4] IRENA. Renewable energy statistics 2023. Abu Dhabi: 2023.
- [5] Khan W, Walker S, Zeiler W. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. Energy 2022;240:122812. <https://doi.org/10.1016/J.ENERGY.2021.122812>.
- [6] Arthur C. What is an energy management system?. United Nations Industrial Development Organization 2021. <https://www.unido.org/stories/what-energy-management-system> (accessed December 4, 2023).
- [7] Graa A, Benhamida F. A review on optimization methods applied to energy management system. Serbian Journal of Management 2020;15:371–82. <https://doi.org/10.5937/SJM15-22519>.
- [8] Maghsoodlou F, Masiello R, Ray T. Energy Management Systems. IEEE Power and Energy Magazine 2004;2:49–57. <https://doi.org/10.1109/MPAE.2004.1338122>.
- [9] Sun H, Guo Q, Zhang B, Wu W, Wang B, Shen X, et al. Integrated Energy Management System: Concept, Design, and Demonstration in China. IEEE Electrification Magazine 2018;6:42–50. <https://doi.org/10.1109/MELE.2018.2816842>.
- [10] Tiwari A, Pindoriya NM. Automated Demand Response in Smart Distribution Grid: A Review on Metering Infrastructure, Communication

Technology and Optimization Models. *Electric Power Systems Research* 2022;206:107835. <https://doi.org/10.1016/J.EPSR.2022.107835>.

[11] Hengrui M, Wang B, Wang H, Wang S, Fuqi M, Wang X, et al. An effective planning approach for integrated energy systems considering equipment operating characteristics. *Heliyon* 2023;9:e21409. <https://doi.org/10.1016/J.HELİYON.2023.E21409>.

[12] Zhao N, Gu W. Low-carbon planning and optimization of the integrated energy system considering lifetime carbon emissions. *Journal of Building Engineering* 2024;82:108178. <https://doi.org/10.1016/J.JOBE.2023.108178>.

[13] Zhou Y, Min C, Wang K, Xie L, Fan Y. Optimization of integrated energy systems considering seasonal thermal energy storage. *J Energy Storage* 2023;71:108094. <https://doi.org/10.1016/J.EST.2023.108094>.

[14] Gao M, Han Z, Zhao B, Li P, Wu D. Optimal planning method of multi-energy storage systems based on the power response analysis in the integrated energy system. *J Energy Storage* 2023;73:109015. <https://doi.org/10.1016/J.EST.2023.109015>.

[15] Guo Y, Xiang Y. IESP: An integrated energy system planning platform. *Energy Reports* 2023;9:242–8. <https://doi.org/10.1016/J.EGYR.2023.04.280>.

[16] Feng B, Fu Y, Huang Q, Ma C, Sun Q, Wennersten R. Multi-objective optimization of an integrated energy system with high proportion of renewable energy under multiple uncertainties. *Energy Reports* 2023;9:695–701. <https://doi.org/10.1016/J.EGYR.2023.05.194>.

[17] He C, Zhu J, Cheung K, Borghetti A, Zhang D, Zhu H. Optimal planning of integrated energy system considering swapping station and carbon capture power system. *Energy Reports* 2023;9:165–70. <https://doi.org/10.1016/J.EGYR.2023.09.120>.

2 Hybrid energy system integration and management for solar energy: A review¹

Abstract

The conventional grid is increasingly integrating renewable energy sources like solar energy to lower carbon emissions and other greenhouse gases. While energy management systems support grid integration by balancing power supply with demand, they are usually either predictive or real-time and therefore unable to utilise the full array of supply and demand responses, limiting grid integration of renewable energy sources. This limitation is overcome by an integrated energy management system. This review examines various concepts related to the integrated energy management system such as the power system configurations it operates in, and the types of supply and demand side responses. These concepts and approaches are particularly relevant for power systems that rely heavily on solar energy and have constraints on energy supply and costs. Building on from there, a comprehensive overview of current research and progress regarding the development of integrated energy management system frameworks, that have both predictive and real-time energy management capabilities, is provided. The potential benefits of an energy management system that integrates solar power forecasting, demand-side management, and supply-side management are explored. Furthermore, design considerations are proposed for creating solar energy forecasting models. The findings from this review have the potential to inform ongoing studies on the design and implementation of integrated energy management system, and their effect on power systems.

¹ This chapter is a paper authored by Tolulope Falope, Liyun Lao, Dawid Hanak, & Da Huo, published in the journal of Energy Conversion and Management X, 2024, <https://doi.org/10.1016/j.ecmx.2024.100527>

Keywords:

Solar Energy Integration, Integrated Energy Management System, Solar Energy Forecasting, Demand Side Management, Supply Side Management, Time-of-Use, Direct Load Control

2.1 Introduction

Currently, about 770 million people globally do not have access to electricity [1]. It is estimated that by 2035, global demand in energy will increase by a third [2]. The upward trend in global energy demand is currently being met mostly by fossil fuels like coal, oil and natural gas, which have as their by-products global warming and emission of greenhouse gases (GHG) [3,4]. In British Petroleum's Statistical Review of World Energy [5], fossil fuel accounted for 82% of primary energy use in 2022, with coal and natural gas growing post-pandemic by 6% and 5.3% respectively. To curb the GHG emissions and meet the emission reduction targets set out in the Paris Agreement, the governments globally have increasingly turned to renewable energy sources (RES) to help satisfy demand. According to BP [5], RES contributed about 10.2% to power generation, reaching double figures for the first time.

RES, like solar and wind, have been widely adapted and are increasingly being used to meet load demand. They have greater penetration due to their availability and potential [6]. As a result, the global installed capacity for photovoltaic (PV) increased to 488 GW in 2018, while the wind turbine capacity reached 564 GW [7]. Solar and wind are classified as variable renewable energy (VRE) and are non-dispatchable due to their intermittent nature [2]. However, economies of scale and improved research and development of VRE technology has led to a decrease in price of VRE components. The combination of price reduction and desire to limit the environmental impacts of fossil fuels has seen a greater penetration of VRE in the grid [8].

Renewable energy technologies increasingly have both off-grid and on-grid applications, but the intermittent nature of certain RES poses a challenge to

grid operators as energy supply does not always match the consumption. As RES penetration levels increase with the advancement of decentralised distributed generation (DDG), adequate steps must be taken to make sure that system constraints like fluctuating supply, capital costs of DDG and grid extensions, and rising load demand, does not compromise the stability of the grid. Grid integration is the process of creating practical, affordable ways to integrate VRE into the grid while preserving or enhancing system stability and reliability. [9].

Solar energy is gaining popularity because of its versatility in various industrial applications including power generation for residential and commercial use, solar drying of fruits for industrial food processing, powering automotives and aeroplanes and so on. Furthermore, governments and organizations are promoting the use of solar energy through various incentives and policies because of its relative safe-use, scalability, and favourable environmental impact compared to other sources. While the installed capacity of wind energy in 2018 was higher than that of solar energy, the latter experienced an annual growth rate of 47.6% compared to 18.9% for the former [7]. In addition, it is projected that around 45% of the world's energy consumption could be met by solar energy by the mid-21st century [10]. In their research, Yao et al. [7] concluded that both demand side and supply side integration costs are lower for solar energy than for wind. PV systems using battery banks are the most popular because they are easier to install and cost less in terms of capital investment than other renewable energy technologies [11]. According to the International Energy Agency's World Energy Outlook 2020 report, solar power was able to achieve the cheapest electricity rate in history. It was cheaper than coal and gas in most countries with price bids as low as 0.0104 US\$/kWh [12]. Combined with increased efficiency of solar cells, and improved economies of scale leading to mass production of high-quality wafers [13], it is easy to see why solar remains a popular choice for power generation. Currently the energy intensity of silicon wafers production is about 6kg CO₂e/kg of silicon metal [14], considerably less than other sources.

An energy management system (EMS) can be used to balance the supply and demand of a power system, which is a key requirement in integrating

intermittent RES like solar energy. However, the emergence of big data, cloud computing, Internet of Things (IoT), advanced metering infrastructure (AMI) and other advances in communication has transformed the conventional grid into a smart grid [15,16]. The technology requirements of the smart grid have necessitated the evolution of the conventional EMS into an integrated energy management system (IEMS) [15]. While EMS can be predictive or in real-time, by definition, an IEMS leverages advancement in technology and communication, integrating predictive and real-time controls to initiate both supply and demand responses in balancing the load and power supply in the grid. While EMS may refer to a single power source, an IEMS has the ability to integrate multiple energy systems and initiate various control strategies.

A comparative assessment of various IEMS architectures, based on the interconnectedness of their individual components is crucial in understanding how they impact the power system, provide grid stability, and ways they can be improved. As can be seen in Table 2-1, the published reviews listed does not give an analysis of IEMS frameworks. For instance, Jafari et al. [17] focuses on the limitations and techno-economic requirements of energy storage systems (ESS). Farag et al. [18] highlights issues with solar integration but focuses more on dust particles. Formolli et al. [19] analyses existing integrated solar energy sites with a focus on recommendations that will improve future urban solar energy projects. Peng et al. [20] focuses on PV generation, ESS and solar thermal for heating. Rubio-Aliaga et al. [21] proposes a multifaceted approach considering economic energy and environment to solve groundwater pumping issues by using RES. In both studies by Ozoegwu et al. [22,23], the focus is more from a policy standpoint.

To address the research gap, which is the absence of a review on IEMS, this paper will review current IEMS frameworks/architectures, identify gaps in current knowledge and propose areas for future research by:

- identifying and providing a comprehensive overview of the IEMS building blocks including energy forecasting, multi-energy generation, energy storage, demand side management and supply side management;

- comprehensively reviewing each building block of the IEMS based on their adaptation and grid deployment as described in current literature; and
- providing a side-by-side comparison of various IEMS architectures.

Table 2-1 Recently published reviews related to solar energy management system and integration

Source	Publication date	Scope
Khan et al. [24]	2022	<ul style="list-style-type: none"> • Summary of various optimization modelling techniques of a hybrid renewable energy system (HRES) • Review of sizing, control and energy management strategies for HRES
Panda et al. [25]	2022	<ul style="list-style-type: none"> • Review of DSM in integrating DER and Energy Storage Systems (ESS)
Rayid et al. [26]	2022	<ul style="list-style-type: none"> • Prospects and challenges of renewable based mini-grid implementation in Bangladesh
Apeh et al. [27]	2022	<ul style="list-style-type: none"> • Review of the contributions of PV to national development
Khan et al. [28]	2022	<ul style="list-style-type: none"> • Review of optimization techniques for HRES couples with Hydrogen technologies
Alami et al. [29]	2022	<ul style="list-style-type: none"> • Review of challenges and solutions to PV technology proliferation
Lamnatou et al. [30]	2022	<ul style="list-style-type: none"> • Review of smart grids /smart technologies related to PV systems, storage, buildings, and the environment
Khosrojerdi et al. [31]	2022	<ul style="list-style-type: none"> • Integrating artificial intelligence in smart grid
Banu et al. [32]	2021	<ul style="list-style-type: none"> • A review on the advancements in methane cracking for hydrogen production
Klass et al. [33]	2021	<ul style="list-style-type: none"> • A review on four ammonia production methods
Ozoegwu et al. [23]	2021	<ul style="list-style-type: none"> • An overview of Nigeria's energy policy objectives and strategies
Stevovic et al. [34]	2021	<ul style="list-style-type: none"> • Using nature-inspired optimization for solar energy integration
Dashtpeyma et al. [35]	2021	<ul style="list-style-type: none"> • Development of a resilient solar-based EMS
Peng et al. [20]	2020	<ul style="list-style-type: none"> • State of the art of solar energy utilization in buildings
Shafiul Alam et al. [36]	2020	<ul style="list-style-type: none"> • Review of challenges and solutions to RES integration
Østergaard et al. [37]	2020	<ul style="list-style-type: none"> • Review of technologies and systems that use RES
Syafaruddin et al. [38]	2020	<ul style="list-style-type: none"> • Modelling and Simulation procedures, control strategy for hybrid power generation
Nižetić et al. [39]	2019	<ul style="list-style-type: none"> • Review of smart technology applications focused on efficiency improvement, sustainable and smart resource management

Azuatalam et al. [40]	2019	<ul style="list-style-type: none"> • Comparison of energy management strategies for small-scale PV-battery systems
Rubio-Aliaga et al. [21]	2019	<ul style="list-style-type: none"> • Multidimensional characterization to evaluate PV integration
Guo et al. [41]	2019	<ul style="list-style-type: none"> • Review green energy integration for wastewater treatment plants
Lauka et al. [42]	2018	<ul style="list-style-type: none"> • Development of a methodology to define solar energy potential for urban planning
Ozoegwu et al. [22]	2017	<ul style="list-style-type: none"> • A review of the past, current, and future status of solar integration in Nigeria.
Khoshkbar-Sadigh et al. [43]	2015	<ul style="list-style-type: none"> • Impact of large-scale PV penetration into the grid
Inman et al. [44]	2013	<ul style="list-style-type: none"> • A review of forecasting methodologies and their applications

Following the introduction, Section 2.2 describes the methods and gives the bases for the various literature reviews that span section 2.3 to 2.8. Section 2.3 presents the power system that is most favourable for the IEMS to operate in. Section 2.4 provides an overview of the state-of-the art in solar energy forecasting and provides a comparison of various forecasting models. Next, Section 2.5 provides a literature review of demand side management and highlights the different types of demand responses. Section 2.6 gives an overview of supply-side management and its application in current literature. Section 2.7 explains how an EMS works, citing examples of its application in literature. Section 2.8 brings together the various concepts from Section 2.3 to Section 2.7 and shows how they combine into an IEMS. A detailed comparison of some existing IEMS frameworks is also presented. Finally, Section 9 is a discussion on the authors perspective of various IEMS frameworks including future research. This concluding section also highlights the key findings and contribution to the IEMS literature.

2.2 Methods

In reviewing the existing literature on IEMS, it was determined that there are five major parts of an IEMS framework that supports solar energy integration: the power system the IEMS operates in, solar energy forecasting (SEF), demand side management (DSM), and supply side management (SSM). David et al. [45] largely substantiates this by conducting a bibliometric study using the Scopus

database from 2000 and 2019 on PV solar energy management. From the analysis, ten potential study areas were established on future research trends including forecasting techniques and DSM, both of which are mentioned as future trends, load status studies, and efficient battery use which has applications in both SSM and DSM, and so on.

2.2.1 Solar energy forecasting

The capacity to reliably estimate generation while keeping some flexibility in demand control will be necessary for solar energy integration into the grid [46]. Effective grid integration and planning depend on improved generation predictions [47]. Energy forecasting is fundamental and crucial to practically every area of the economy, including residential, commercial, and industrial [48]. There are research methods that focus on predicting load consumption patterns or peak demand as a way of renewable energy integration [49–52]. However, for an energy system that uses solar energy as its primary source of power, the uncertainty in the system lies more in the variable nature of solar energy. In addition, many loads are invariably affected by weather parameters like solar radiation [53,54] which makes load consumption the dependent variable. Therefore, it is more effective for the stability of a solar-driven energy system and the dispatch of solar energy to the grid, to accurately predict solar energy supply than load consumption.

To support the theory above, Cai et al. [51] concludes that the prediction of energy consumption has to do with improving grid quality and allocation of power supply. Basmadijan et al. [46] on the other hand states that maintaining an equilibrium between power demand and supply solely rests with the supply side. For efficient MG sites that have PV generation units, a precise characterization of the solar energy available is crucial [55]. The argument is that supply is more easily controllable than demand because of the stochastic or unpredictable way consumers use energy [56]. For conventional grid sources, EMS is based on demand forecasting and supply planning. For RES, EMS is based on supply forecasting and demand planning. Therefore, the answer to the question of

renewable energy integration is more suited to the prediction of renewable energy supply also known as solar energy forecasting (SEF).

By being able to accurately forecast solar resources, one can then predict what generation will be like and reduce uncertainty [3,47]. Accurate energy supply forecast informs what control measures can be taken to maintain energy supply-demand balance in the system. For instance, in periods of low energy supply, alternative or standby generators can be deployed to augment power generation. Operational control of both supply and demand in a power system is crucial to renewable energy integration and aids in the balance of power [3].

2.2.2 Energy management system

With the emergence of solar energy as an important source of energy supply, more thought must be given to increasing its usage in an efficient and sustainable way. One such solution is the use of an EMS to match demand and supply. An EMS is a tool that combines complex software and hardware to monitor, control and optimize energy use to reduce cost [52]. The EMS has a control system program that initiates a response depending on the energy supply forecasted. This response could either be a supply response (SR), demand response (DR) or a combination of both. Control systems are an important counter to the fluctuating and intermittent nature of RES like solar and wind energy [57]. An electric power control system uses control loop mechanisms to manage, regulate and direct the electrical components within a power system, and thus the power system itself [57]. Control systems use a feedback controller to modulate control. Parameters such as system frequency or voltage could be used as the process variables where a pre-determined control signal is generated once there is a difference between the value of the process variable and a reference value.

2.2.3 Supply side management

EMS using control systems is critical in implementing SR, DR or a combination of both, to ensure sufficient energy supply to satisfy demand [58]. SR is a subset of SSM. SSM refers to the efficient generation, transmission, and distribution of electricity to meet customer demand [59]. SSM is crucial in creating an energy

reservoir to supplement inconsistent energy supply resulting from intermittency of some RES. There are several SSM techniques including integrating battery storage [60], optimizing and synchronizing alternative or back-up power [61], dispatching algorithm for multiple energy supply sources [62], ramping up and down of generation through automatic control and so on.

2.2.4 Demand Side Management

Conversely, DR is a subset of DSM. DSM is the process of using energy efficiently by managing customers' energy consumption [63], in addition to reducing energy costs and emissions [64]. In the resource-constrained areas of the grid, usually found in off-grid rural communities, storage batteries are often over-sized to increase the reliability of the power system. DSM can act as an energy buffer while reducing the capital and operating costs of the power system. Examples of DSM are energy storage using, energy efficiency in buildings [65], initiating DR and so on. DR focuses on short term adjustments made to load usage [60] and is broadly categorized into two: price-based DR and incentive-based DR. For efficient energy utilization, there is a need to integrate RE and other distribution energy resources with DSM [66]. Conversely, successful implementation of DSM can be achieved through energy forecasting [67]. In addition, efficient power scheduling should be done using RE forecasting and managing electricity tariffs through DSM [68]. A reliable EMS is vital to the success of DSM [69]. It is clear to see that SEF, SSM and DSM are strongly linked in an EMS.

In general, the review found that scientific journals mostly addressed optimising the system in terms of storage, energy dispatch/supply, consumption, and equipment sizing. Largely undiscussed was a comparative overview of the IEMS framework. To analyse IEMS frameworks, we conducted a comprehensive literature review for each sub-section of the framework, including the EMS. The aim was to identify and emphasize key concepts and design considerations for each parameter and their interconnectedness within the IEMS. Figure 2-1 illustrates the approach used in the study.

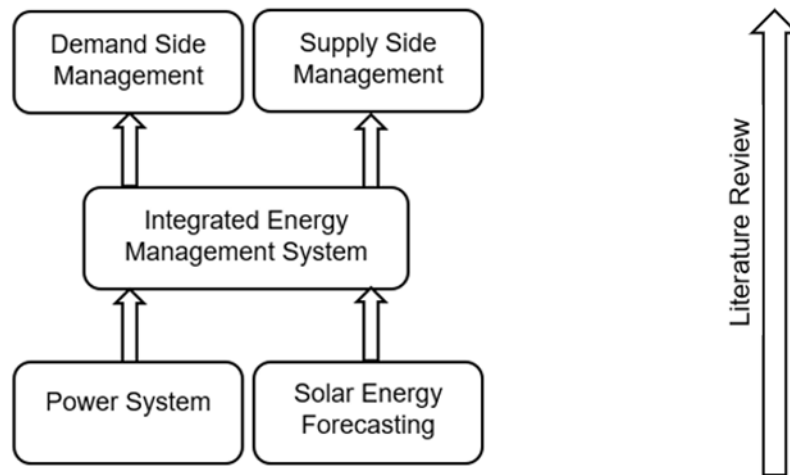


Figure 2-1 Review approach for IEMS Framework

2.3 Renewable smart hybrid mini-grids: The future of renewable energy integration

A smart MG is a collection of controllable and physically proximate distributed generator(s) and load resources, where there are multiple sources of AC power and at least one of these is based on a renewable energy technology such as wind or solar energy [4]. With the advent of smart grids, these resources can be better controlled to deliver efficient and reliable power despite their intermittent nature [46]. Due to their modular and distributed nature, smart grids are a viable and sustainable way to provide power in developing countries [70]. The renewable smart hybrid MG is ideal for dispatching an EMS because it requires smart communication of multiple load and supply sources at the same time. There are different classifications and iterations of the smart grid or MG namely the renewable energy home system (REHS), renewable hybrid mini-grid (RHMG) and the renewable smart hybrid mini-grid (RSHMG). Figure 2-2 shows the three different configurations. An examination of their different characteristics and their importance to the deployment of an EMS is further explained.

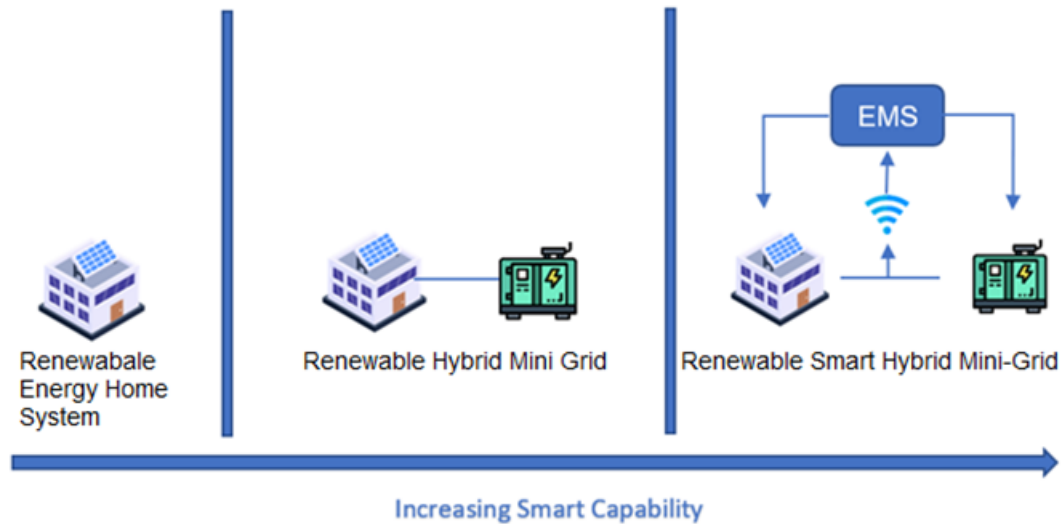


Figure 2-2 Power system configurations for energy management system deployment

2.3.1 Distributed generation: Mini-grid & renewable energy home systems

The conventional method of providing new electricity connections mainly from traditional sources like fossil fuels is not sustainable, as the environmental costs and rising fuel prices make grid extension expensive and less attractive [4,71]. For the problem of universal electricity access, the idea of a central utility gradually extending the grid is being overshadowed by a more robust solution using DDG. While the former causes transmission losses [72], the latter provides the use of locally available energy sources. This eliminates transmission costs and reduces power losses [4] as well as increases the flexibility and reliability of the overall system.

DDG, particularly using RES, has proven to be a viable answer in terms of cost and flexibility [3]. DDG shows a greater capacity of renewable integration without compromising the grid [73]. DDG renewable configurations differ with respect to capacity. At one end of the spectrum are MGs and integrated solar farms, while at the other end are smaller renewable energy home systems (REHS) like rooftop solar arrays. The increase in solar energy penetration has also largely increased due to the fact that RES now have small-scale and large-scale applications [74] with users ranging from small residential customers to large utility operators. This is not independent of the fact that RE is relatively

affordable and can be accessed by a wide spectrum of customers [2]. Figure 2-3 gives an overview of the DDG classifications.

The Nigerian Electricity Regulatory Commission (NERC) defines MGs as systems with integrated generation and distribution networks that have installed capacities below 1 MW [75]. This can be with or without storage [76]. NERC also goes further to classify MGs as being either isolated (no connections to any network) or interconnected to the main grid [75]. A study by the International Renewable Agency (IRENA) gives a more general definition of MGs as simply an integrated structure of energy sources and loads that may or may not include functions of generation, storage control, management, measurement, conversion, and consumption [77]. MGs are increasingly popular because of technology flexibility and an efficiency in integrating large scale DDGs [78].

REHS are gaining popularity due to lower installation and operational costs [11]. Like the name implies, these are systems that use one or more RES as its primary or back-up source. They can either be stand-alone or grid-connected. An example of a REHS is the solar home system (SHS). The SHS is comprised of solar panel(s), a charge controller, battery, an inverter, and a load. At the very basic level, the inverter is not included, and the load is DC in nature. REHS can range from several watts to several thousand watts. They usually represent the low end of the DDG spectrum.

REHS and MGs are quite similar in the sense that they both operate a decentralized grid. In fact, MGs may be viewed as a bigger version of the REHS in terms of energy production capacity. Ekpe and Umoh [70] infer that a solar home system (SHS), which is an example of a REHS, can be easily scaled up to a MG. There are, however, some other differences. REHS tend to provide a lower tier of service than MGs [77]. Though central utilities on an average provide a higher tier of power supply to multiple customers than both, they must consider power losses due to distance and its equivalent costs. MGs on the other hand can provide the quality of service a central utility provides while being closer to the customer. In fact, MGs are a replica of the much larger central power systems albeit on a smaller scale [79]. REHS are tailored to a particular customer's load

profile [76] while MGs can reasonably be designed and optimized to match the aggregate load profiles of the customers it supplies [80].

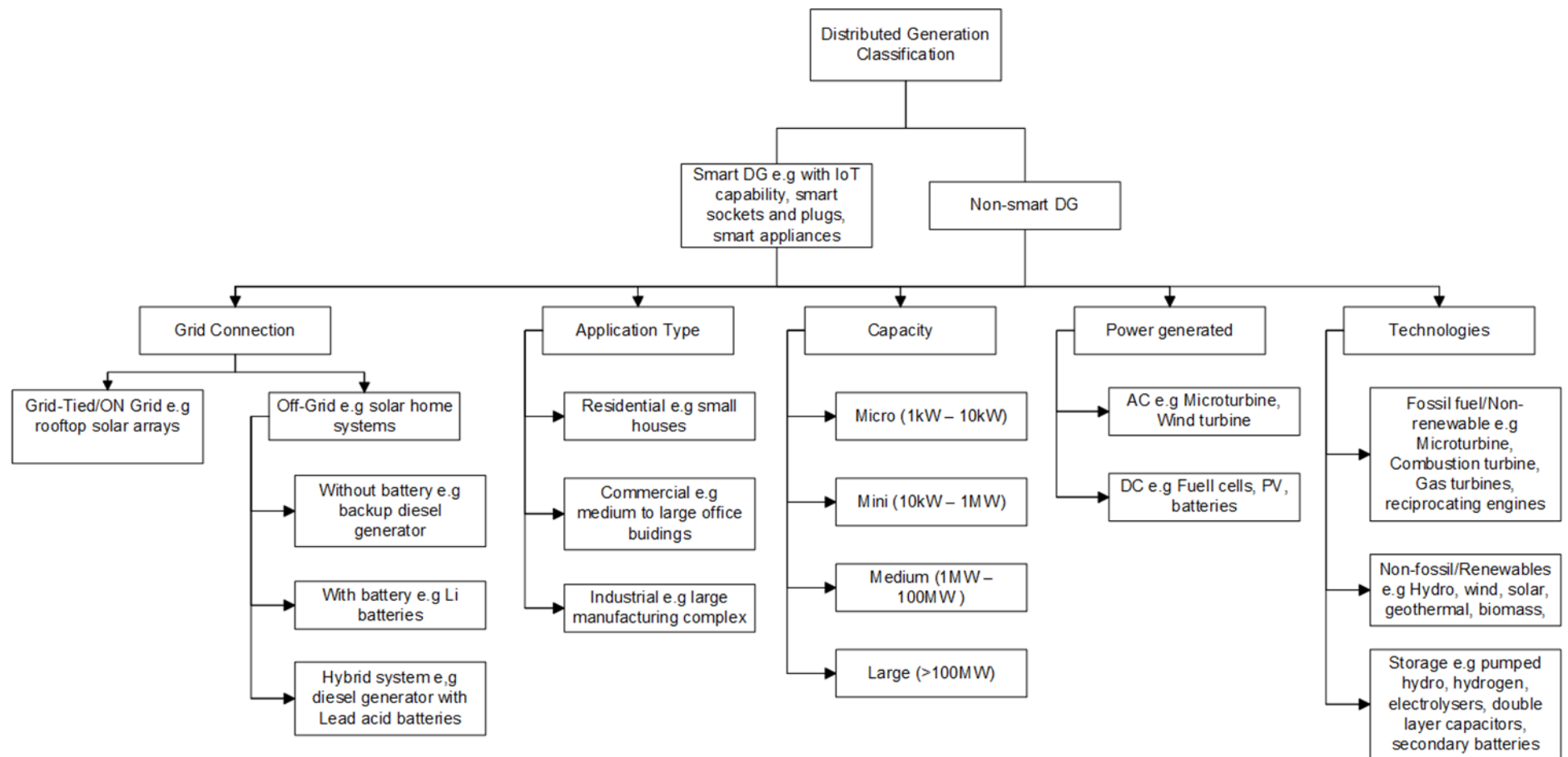


Figure 2-3 Classification of distributed generation system

2.3.2 Renewable mini-grids and hybrids

A MG based on some-form of renewable energy is referred to as a renewable mini-grid (RMG) [77]. RMGs are viable alternatives in supplying areas with no connection to the grid (off-grid), areas with inefficient or erratic energy supply, or areas where grid extension is not cost effective [3,77]. These MGs can serve as a backup to the main grid or at other times help supply power during peak demands [63], thereby reducing the burden on centralised utility networks.

RMGs can also be paired with generation from non-renewable sources like diesel generators to form a renewable hybrid mini-grid (RHMG) [63,70,77]. RHMG are a cost-effective way of improving power system reliability through renewable energy integration [81,82]. Typical configurations of the RHMG have the renewable component as the primary source while diesel and/or batteries serve as backup [83]. RHMGs are able to combine the advantages of each generation source. For instance, a solar-diesel MG takes advantage of the stable power supplied by a diesel generator versus the intermittent nature of solar energy, while solar as a renewable energy source isn't affected by varying fuel prices. As compared to single-source power systems, RHMG supply greater load for longer and are more efficient and cost-effective [84]. Falk et al. [85] also attributes faster and more flexible energy supply as advantages of RHMGs especially when incorporated with some form of storage. RMGs are viable in providing reliable, affordable and environmentally friendly supply of energy [63,77,80].

2.3.3 Grid-tied versus off-grid

IRENA groups MGs based on their connection to the grid and the type of service provided. Connections are categorized as autonomous (isolated) or interconnected (grid-tied) while the type of service is either lower tier or higher tier [77]. However, Szewczuk [86] makes the case for an interconnected system over an isolated system by highlighting the lower cost per connection due to the fact that power from the grid can sometimes substitute storage and help with system balancing. In addition, since grid-tied MGs are part of the main grid, they

can be configured to help with the utility's short-term problem of on-grid congestion and peaking loads [63], and its long-term problem of grid extension. Furthermore, Robert et al. [79] sites the semi-autonomous design of the interconnected MG as the reason it contributes to the overall reliability and resilience of the system. Isolated or off-grid MGs can only help with the latter problem and require complex energy balancing mechanisms if the primary source is RE. There is, therefore, a stronger case to look at an interconnected system that addresses both on-grid congestion and grid extension. Typical configurations of the RHMG are operating:

- as the primary power provider with the main grid as a back-up;
- as a back-up to the main grid; and/or
- a system to shave up peak demands and supply critical loads [77].

The EMS should be robust enough to be deployed either for on-grid or off-grid power systems. The EMS's function for an on-grid application will be to reduce the strain on the grid by optimising the operations of any energy storages, co-ordinate power dispatch and reducing peak load demands. In an off-grid application, the EMS's main function will be to maintain the integrity of the grid by matching supply to demand since there is no back-up as the main grid is absent. EMS can also be applied to frequency and voltage control of the grid.

2.3.4 Renewable smart hybrid mini-grid

Deployment of interconnected MGs has been limited and the technology is still emerging [77]. The ability for various equipment and components in the power system to interact and communicate with each other, interpret, and respond to external and internal inputs in order to preserve the stability of the system makes the grid "smart". The objective of the smart grid is to upgrade the existing power system to one that is safe, adaptable, resilient, expandable, and sustainable [87,88]. Adding intelligence or smarts to a system yields better results for all stakeholders involved, for example optimally switching power sources and load shedding [3]. To increase grid penetration, RMGs or RHMGs "require technology advancements in design and planning phases" [77]. According to IRENA [77], these technological advancements include intelligent controls that can interpret

high level algorithms, and integrate accurate RES predictions with adequate battery control. Moreno-Garcia et al and Judge et al. [89,90] identify these technologies as AMI, smart meters (SM), sophisticated communication and information architectures or advance control and automation techniques. The incorporation of these technologies transforms the RHMG into a renewable smart hybrid mini-grid (RSHMG) [63]. Palahalli et al. [91] describes smart grids as the interaction between power and communication devices to control power flow within the grid network. These communication devices must be reliable [70] so that the flow of information between monitoring and control devices results in the accurate response. Smart grids must be able to integrate all components of the system, from supplier to the end-user, and possess real time monitoring and control [89]. The emphasis on smart grid is because the central technology used in implementing DR programs or cost-reflective pricing (CRP) is the SM [92]. SM is part of the AMI that integrates measurement, communication, control and monitoring between the supply and demand side of the system [90,93].

Smart grids have also given rise to smart homes. Smart homes with residential power generators are also capable of monitoring and controlling customer appliances and loads [65].

2.4 Solar energy forecasting

The push for integrated renewable energy generation is seen as a key step in reducing the dependency on depleting fossil fuels used in power generation. However, the intermittent nature of RES, like wind and solar energy, means that a higher penetration of these sources in the traditional grid would lead to reliability and quality issues [13,79,89,94]. Therefore, a reliable way to forecast energy resource availability is crucial in making sure energy demand always matches supply. Accurate resource forecasting reduces the need for storage and other reserves [79] and helps the energy sector to minimize power fluctuations while maintaining the overall reliability of the system [66]. The results of a study by Kromer et al. [95] show that by incorporating forecasting and day-ahead shaping

of customer's load, need for storage was reduced by 10%-20% and levelized cost of energy (LCoE) by about 10%.

SEF is the process of predicting future solar irradiance or solar power generated from historical and/or present meteorological observations [96]. There are many SEF methods and techniques usually derived from considerations like data input, forecasting architecture used to map the input data, the forecasting methodology, the forecasting time frame/scale and the predicted outcome [97]. In considering the data type, this could be historical, real-time, or forecast data. The data source describes the origin of the data for example if data comes from an on-site weather station as opposed to a local weather station. Table 2-2 shows a review of some studies based on data source and types.

Table 2-2 Review of research studies on data source and types

Research Study	Historical Observed Weather Data from Weather Station	Forecast Weather Data	Historical Observed Weather Data On-Site	Historical Observed Energy Meter Data
Sarp et al. 2021 [98]			✓	✓
Aprillia et al. 2020 [99]			✓	✓
Cannizzaro et al. 2021 [100]			✓	✓
Gu et al. 2021 [101]			✓	✓
Nguyen et al. 2021 [102]			✓	✓
Pan et al. 2021 [103]			✓	✓
Ngoc-Lan Huynh et al. 2021 [104]	✓			✓
Pedregal et al. 2021 [105]	✓			✓
Ahmad et al. 2021 [106]	✓			✓
Korkmaz 2021 [107]	✓			✓
Li et al. 2019 [108]	✓			✓
Najibi et al. 2021 [109]	✓			✓
Rodríguez et al. 2018 [110]	✓			✓
Thukral 2020 [111]	✓			✓
Wang et al. 2018 [112]	✓			✓
Yakoubi et al. 2021 [113]	✓			✓
Zang et al. 2018 [114]	✓			✓
Chen et al. 2020 [115]	✓			✓
Anaadumba et al. 2021 [116]				✓
Kushwaha et al. 2019 [117]				✓
Iyengar et al. 2014 [118]		✓		✓
Andrade et al. 2017 [119]		✓		✓

Leva et al. 2017 [120]		✓		✓
Persson et al. 2017 [121]		✓		✓
Abedinia et al. 2017 [122]	✓	✓		✓
Carrera et al. 2020 [123]	✓	✓		✓
Kim et al. 2019 [124]	✓	✓		✓
Kyliashkina et al. 2019 [125]	✓	✓		✓

Table 2-2 shows the possibility of different permutations of data types and sources that have not been explored. There is clearly room for further research and contribution to this area as highlighted by David et al. [45] in Section 2.1. While some studies may choose to use one data (univariate) type, Table 2-2 reveals that majority rely on two data sets: the historical PV generation and a meteorological data type. Even fewer studies make use of more than two data sources. A more in-depth analysis of current literature, and an expansion of the concepts highlighted in Table 2-2, can be seen in Table 2-3.

Table 2-3 A summary of studies on solar energy forecasting

Research	Power System	Data Source	No. of Weather Data Parameters O - Observed F - Forecast	On site Weather Station	Forecasting Methodology	Forecasting Architecture	Prediction time	Predicted Outcome
[123]	93MW hybrid PV-wind	* Forecast weather * Historical weather * PV data * Sun elevation	f - 7 O - 16	No	* Deep feedforward network (DFN) * Recurrent neural network (RNN) * Hybrid network (HN)	* Forecast → DFN * Observed → RNN * DFN + RNN → HN = solar energy	24-hour-ahead	Solar power generation
[126]	74kW DC Microgrid	* Historical weather * Real time weather in 30 min intervals * Time * PV Data * Wind Data	O - 5 Real time - 3	Yes	Support Vector Machine (SVM)	* Historical + real time → SVM = solar irradiation * Solar irradiation → solar cell equation = solar energy * Historical → SVM = wind speed * Wind speed → wind energy generator equation = wind energy	24-hour-ahead	* Solar power generation * Wind power generation
[65]	* 2.3kW Wind Turbine * 3.3kW Solar System * 22kWh battery storage	* Observed weather * PV Data * Wind Data	O - 4	Yes	* Wavelet Transform (WT) - 3-level Wavelet decomposition * Artificial Neural Network-multilayer feed forward back propagation (FFBP) network	* Wind speed → wind curve = wind power generation via wind curve * Solar radiation → I _{sc} , V _{oc} formula = solar power generation	5-min-ahead	* Solar power generation * Wind power generation

[62]	Hybrid Microgrid * 25kW Solar System * 15kW Wind Turbine * 300kW Diesel * 30kW Fuel Cell * 30kW Battery * 30kW Grid * 30kW Water Micro Turbine	* Forecast weather * PV Data * Wind Data	f - 2	No	* Probability Density Functions (PDF) - Solar * Rayleigh distribution - Wind	* Forecast → PDF = solar irradiation * Solar irradiation → solar power output formula = solar generation * Forecast → Rayleigh = wind speed * Wind speed → wind generation formula = wind power	24-hour-ahead	* Wind Speed * Solar irradiation
[127]	* Solar PV system * Micro-hydro turbine power system	* Current - Imp * Voltage - Vmp	N/a	No	Artificial Neural network: Feed-forward with Levenberg - Marquardt back propagation	Imp + Vmp → feed-forward = solar generation	24-hour-ahead	Solar power generation
[128]	640W PV system	* Forecast weather * Historical weather * PV Data	N/a	No	Random Forest: Non-linear multiparameter regressor	Forecast + historical → Random Forest = solar generation	24-hour-ahead	Solar power generation
[66]	N/a	* Historical weather * PV data	O - 5	No	Particle Swarm Optimization (PSO) based Support Vector Machine (SVM) regression model	Historical → PSO with SVM = solar irradiation	24-hour-ahead	Solar irradiation

[11]	14.52kW PV	* PV data * Load demand data	N/a	No	Residual Dilated Causal Convolutional Network (Res-DCCN), Naïve, Support Vector Machine (SVM), Artificial Neural Network (ANN)	PV data→ Res-DCCN-A = Solar generation forecast Load data→ Res-DCCN-B = Load demand forecast Losses from Res-DCCN-A+Losses from Res-DCCN-B = Dual modal loss→ Adam Optimizer	1-hour-ahead	Solar power generation
[129]	N/a	* PV data * Wind data * Load demand data	N/a	No	Probability Density Function (PDF) based on kernel method using Monte Carlo Simulation	Historical → PDF using Kernel function and density estimator → Monte Carlo simulations = Solar irradiation + wind speed	N/a	* Wind Speed * Solar irradiation
[13]	5 kW PV	* Historical weather * PV data	O - 4	No	Wavelet decomposition-Generalized Neural Network (GNN) based online forecasting model	N/a	15-min-ahead	Solar irradiation
[130]	N/a	* Historical weather * PV data	O - 4	No	K-nearest neighbour (KNN)	*Historical data → Similar analysis using KNN *Weighed average of similar days = solar generation forecast	24-hour-ahead	Solar power generation
[131]	434kW PV	* Historical weather * PV data	O - 1	N/a	Modified Support Vector Machine (SVM) with Gauss Newton Method nonlinear least squares	Historical solar insolation→ Modified SVM = solar generation	24-hour-ahead	Solar power generation
[132]	*1.4kW PV * Grid Connected	* Historical weather * PV power	O - 2	Yes	* Principal Component Analysis (PCA)	*Historical data → PCA *Historical data → FRNN	N/a	Solar power generation, current and voltage

					* Full Recurrent Neural Network (FRNN)			
[133]	N/a	Historical weather	O - 11	No	* Random Forest (RF) * Artificial Neural Network (ANN) * XGBoost	*Historical data → PCA +feature *cleaned data→RF,ANN,XGBoost	N/a	Solar irradiation

In Table 2-3, the second column describes the type of power system used in the forecast. As can be inferred from Section 2.2, each power system is unique and has characteristics that will influence how the solar energy is forecasted. In the data source column, data used in the studies were historical and forecast weather data as well as power or energy meter readings from the PV plant. Some studies use more weather data parameters than others in selecting the predictors to train the forecasting model. These predictors have a direct impact on the predicted outcome. This is important because the more relevant data used in training, the more accurate the model is [110]. In summary, the more weather parameters considered, the higher the possibility of getting better predictors.

The proximity of the weather data to the power system is another factor to consider when assessing a forecasting approach. Although weather data from a local weather station may be accurate, it is possible that they don't fully reflect the site's local characteristics. By having a weather station on-site, this weather information can be further refined [2]. Local weather conditions that affect the weather but are not represented in data from local or regional weather stations can be captured by the on-site weather station. Therefore, the presence of an on-site weather station is important and is included in column 5 of the table.

According to the temporal scale, solar forecasting is divided into ultra-short-term (1 minute to 1 hour ahead), short-term (1 hour to 1 week ahead), medium-term (weeks, months, and quarters) and long-term forecasting (1 year to several years ahead) [134].

The forecasting architecture describes how the forecasting methodology maps the input data to the predicted outcome. Solar forecasting methods are divided into physical, statistical, and hybrid models [135]. Physical methods are divided into Satellite-Imaging models, Sky Images, and Numerical Weather Predictions (NWP) [136]. NWP uses meteorological weather data like pressure, temperature, humidity and so on, to predict either the solar radiation or PV generated. NWP output data are the most widely used for 24-hour or day-ahead forecasts [137,138]. Statistical methods like Artificial Neural Networks (ANN) and Support Vector Machines (SVM), are data-driven and rely on historical data to make predictions. Both the physical and

statistical models can be combined to form hybrid models that provide a higher forecasting accuracy.

2.5 Demand side management

Power system management can be categorized into demand side management (DSM) and supply side management (SSM) [139]. Increase in energy demand and prices necessitates energy optimization at both the supply and demand side [65]. SSM and DSM are both critical in the planning phase for the integration of intermittent RE [140] and are useful in reducing peak loads and increasing network capacity [139]. Energy planning must consider the effects of both DSM and SSM in designing energy systems that are reliable, financially viable and environmentally responsible [59]. The ultimate goal of integrating both DSM and SSM is to meet customers demand needs in a reliable and cost-effective way.

DSM and SSM can be uniquely viewed with regards to ramping up or down the demand or supply respectively. In terms of DSM, the goal is to reduce consumer demand to a manageable level at which load demands can be effectively supplied by the system. On the other hand, DSM can be used to increase demand by activating shiftable demand load and storage during times of surplus energy production.

As described by Szewczuk [86], smart grids make use of smart technologies that focus on concepts like “*dynamic demand management, automated battery control, low-cost solar forecasting, intelligent refrigeration, optimal grid planning and automated fault detection and diagnosis*”. Kakran et al. [63] include components like demand side responses and other smart devices like smart meters. The purpose is for increased grid reliability, significant savings, and improved power quality [71].

DSM is the process of using energy efficiently by managing customers’ energy consumption [63], thereby reducing the load demand and maximizing the capacity of the power system. It refers to monitoring and control processes that reduces the energy demand at the customer’s side of the meter [141]. Despite the advantages of the RSHMG, the intermittent nature of renewable energy sources like solar energy can lead to periods of reduced electricity generation. Also, non-existent or constrained storage, along with fluctuating peak demand, can often reduce the reliability and

efficiency of the system [142]. While energy storage can help mitigate some of these problems, its relatively high cost hampers its deployment [67,143]. At these times, DSM can be a useful tool to reduce demand peaks [93] and more evenly spread demand so that it better matches with supply [142,144]. DSM is also listed as a key resource that can help maintain stability in the grid when supply and demand do not match up [13]. With no control over the supply, the aim is to modify the load to fit the supply pattern [64]. In summary, DSM integrates load shifting and energy efficiency to balance supply with demand [145]. Harper lists price incentives and distributed intelligent load controllers as examples of DSM techniques and DSM technologies, respectively. DSM implementation can potentially reduce the total investment costs of MGs [67]. For example, remote or off-grid battery system storage can be reduced as DSM can account for times of high peak. In an analysis of a SHS by Mehra et al. [146], by grouping the load into critical and non-critical (shiftable) loads, the capital cost of the system was reduced by 26%. In terms of cost to the investor/operator, DSM can move demand to periods of excess generation, thereby eliminating or limiting the need for energy storage [143]. For consumers, DSM means cheaper electricity through effective and optimal load scheduling based on current electricity prices. The implicit benefits of reducing peak demand include mitigating against electrical system emergencies, deferring the cost of building additional transmission and distribution networks, and reducing blackouts [59].

DSM is also critical to stand-alone systems since backup power cannot be called from the grid [126]. This is even more pronounced if a RES is used as the primary source. The DSM architecture [141] is shown Figure 2-4.

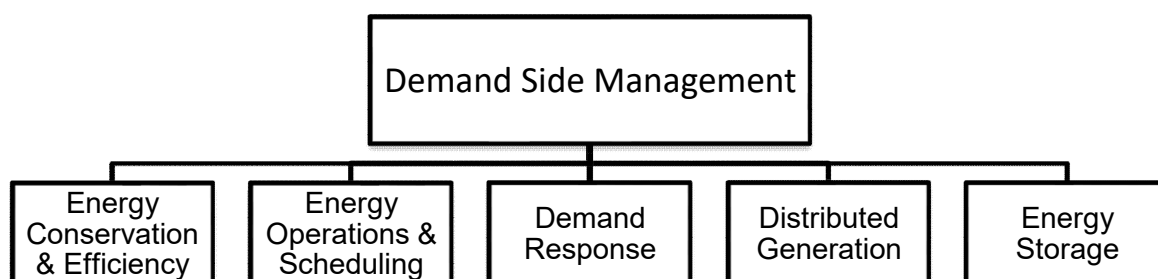


Figure 2-4 Demand side management architecture

2.5.1 Demand response

DR is a process whereby the energy demand of a customer is reduced by directly or indirectly shutting off or reducing the consumption of some of the customer's appliances [141]. In effect, customer loads are shifted from peak hours to off-peak hours [90,147] or from periods of high generation to lower generation [13]. Parrish et al. [148] provided a broader definition of DR to include the flexibility to increase as well as decrease energy demand to respond to surplus energy or reduced demand peaks, respectively. When the energy supply is low, non-essential loads can be turned off and later re-scheduled to periods of higher generation. During peak load periods, a utility can shave off load by incentivizing customers to turn off equipment and shift loads to off-peak periods. The incentive must be structured so that it provides participating customers with payments exceeding their operational value for short durations a few times a year. Customers can either pay higher tariffs during peak periods or move non-essential load to off-peak periods and still get a discount.

The flexibility of DR is well suited to meet the fluctuating supply levels of variable RES [8] and is an emerging tool at very high penetrations of VRE [9]. It is also important to note that DR focuses on short-term adjustments made to load usage [60] as a way of balancing the energy in the system rather than a more permanent long-term fix as suggested by many SR methods. DR is often referred to as a bottom-up approach, where the flexibility of the customer's demand cancels out the variability of RES, making the system more reliable and stable [149]. DR can be categorized into two [141] as shown in Figure 2-5.

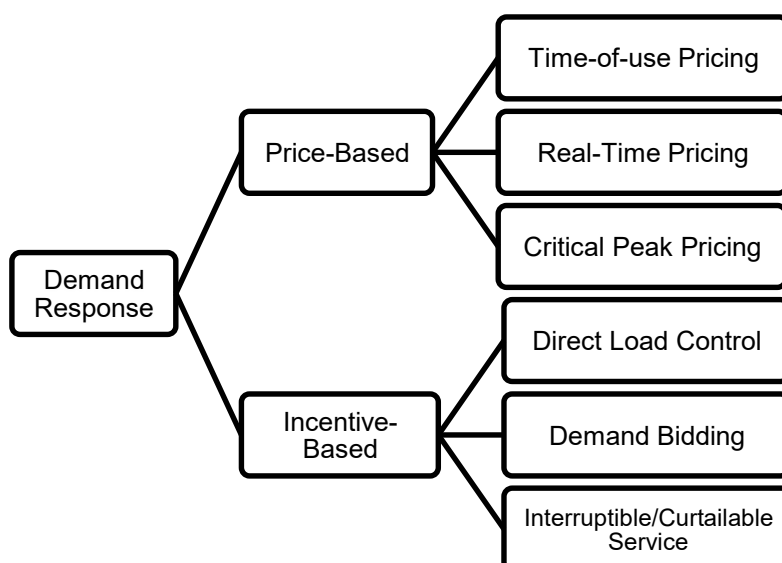


Figure 2-5. Demand response architecture

Typically loads in the DR program are categorized into interruptible, schedulable, and non-schedulable loads. Interruptible loads are loads that can be turned off for short periods without too much discomfort to the consumer [65] for example, water heaters. Examples of schedulable loads are those that can be controlled by a thermostat, such as air conditioners and space heaters, as well as those that don't necessarily need to run at a certain time, including dishwashers, clothes washers, and dryers. [8]. Minhas et al. [126] classified these loads as “*controllable or flexible load demands*” examples of which are refrigerators and charging electric vehicles. Non-schedulable or non-controllable loads are sometimes referred to as base loads [126]. These loads are of utmost priority to the customer, for example, light sources and power sockets. Some other literature may use different nomenclature like shiftable and uncontrollable loads for schedulable and non-schedulable, respectively [150]. It is important to note that while some loads like water heaters may fall under both interruptible and schedulable loads, other loads like washing machines, electric cookers and dishwashers should not be interrupted during operations. Therefore, schedulable tends to refer to these types of loads.

An alternative to the direct participation of users, sometimes referred to as customer control, is the process of demand response automation using existing control systems [79]. The latter is also known as direct load control, which may use pre-agreed load profiles or dynamic real-time profiling according to hour-ahead forecasts to shut down and re-start certain customer load appliances.

DR programs are also linked to customer perception and satisfaction. The ability of customers to reduce costs by scheduling their load based on preference is a key motivation for customer participation in DR programs. DR also create “virtual” power that can be called upon to reduce the load. By optimizing DR, system operators can build smaller MGs and reduce capital and operating costs.

2.5.2 Price-based demand response program

Refers to changes in power usage by customers in response to changes in electricity prices in the form of tariffs. These changes would occur if the corresponding tariff associated with different periods is significant enough. In other words, customers are more likely to change their consumption habits if they are required to pay for more during peak hours as opposed to off-peak hours [59]. Price-Based DR programs can be categorized into Time-of-Use (TOU) pricing, Critical Peak Pricing and Real-Time Pricing [141].

2.5.2.1 Time-of-use pricing

TOU pricing is based on the time interval for which the electricity is used [93]. Typically, a day is divided into three intervals: peak interval, mid-peak interval and off-peak intervals [93] with some literature referring to the mid-peak as the “shoulder” interval [68,151]. The charges for operating loads during peak periods are usually significantly higher than at other time intervals, in many cases double or quadrupled [145]. In this way, the consumer is encouraged to either reduce peak period loads or shift them completely. Using less energy during peak periods and more during off-peak simulates the effect of peak shaving and valley filling, leading to a more balanced energy system [150]. For example, flexible loads, such as preheating electric hot water heaters or charging electric cars, are shifted to times of high solar energy production where the cost is likely to be cheaper than the traditional grid electricity. Other time frames can be time of day (day versus night), day of the week (weekday versus weekends) or even time in the season (summer vs winter) [152].

TOU can be further broken down into static or dynamic. A static TOU tariff offers a minimum of two different rates for fixed periods in the day. For example, an operator can charge higher rates for weekday evenings and a lower rate for mornings. The dynamic TOU tariff offers different rates for different hours of the day, usually with input from real-time load/supply forecasting. It is also important to note that both TOU rates can either be automated or involve direct customer participation.

TOU is considered one of the main pricing strategies used in the power market, along with RTP and multistep electricity pricing [49]. It has wide applications in the electricity markets because it has many advantages [13].

2.5.2.2 Critical peak pricing

Critical peak pricing (CPP) are tariffs based on the peak hours of the day. Peaks are sometimes determined by the real-time energy market, or if a contingency or an event such as equipment breakdown, or high temperatures, occurs [153,154]. The aim is to reduce the short-term peak in consumption. Peak hours logically will have a significantly higher energy cost than off-peak periods to incentivize customers to operate their loads during the latter [155].

2.5.2.3 Real-time pricing

Real-time pricing (RTP) tariffs are charged over very small intervals to reflect the true nature of fluctuating energy prices. In other words, this tariffs frequently correspond to wholesale market prices [156]. RTP, also called dynamic pricing, mirrors a utility's production cost in real time [157]. To capture such granular data requires relatively expensive data monitoring and collection technology, high rate of data collection and efficient data processing techniques. This makes implementation of RTP expensive than other tariffs.

2.5.3 Incentive-based demand response program

Incentive-based DR programmes are usually established by the utility or operator where the customer is given added financial incentive to reduce their load during peak periods [93] or during periods when the utility/operator determines that not reducing demand will lead to a system collapse.

2.5.3.1 Direct load control

Direct Load Control (DLC) is the process whereby the demand response operator remotely shuts down or reduces a customer's load or appliance [93,158,159]. The choice of customer loads are loads that can bare such short-term interruptions [159] with little inconvenience to the customer. The decision to shut down certain loads is usually taken by the system operator to preserve the stability of the grid (sometimes called load shedding) or by the customer to reduce operational costs. A number of studies have evaluated DLC. Parrish et al. [160] used DLC to shed unnecessary loads

by using undervoltage and underfrequency relays to monitor customer loads. This method was shown to have up to 97% power supply reliability.

Some DLC programs use load controllers or load relays [159] that may be centralized [151] to achieve instant load control. Customers are offered discounted rates or incentivized payments to participate in such programs. Acute circumstances that threaten the integrity of the grid tend to be the reasons for activating DLC [161]. This means that the customers may not be pre-notified, nor consent sought due to the urgency of the situation.

2.5.3.2 Demand bidding

Demand bidding or demand side bidding is a concept that enables consumers to participate directly or indirectly in electricity trading, with favourable tariffs or cash incentives often tied to some demand management. In other words, operators can offer customers cash incentive to achieve a daily load reduction target [162]. Since most customers are not experts in the nuances of the electricity markets, they may choose to trade indirectly through electricity retailers or municipalities who act as intermediaries between the market and the customers [163]. These intermediaries submit bids on behalf of the customers and, through their expertise can negotiate cheaper rates based on their knowledge of electricity prices [164].

2.5.3.3 Interruptible/curtailable program

In a summary, interruptible or curtailable programs are those that use financial incentives to encourage load-shedding responses [165]. They take advantage of a customer's interruptible loads by switching them off or keeping them on pause, during peak or emergency events to maintain the required reserve capacity and preserve the stability of the system [165]. This is achieved when the system operator looking to obtain reserves, signs an agreement with the customer to interrupt their load based on some agreed pre-conditions. In return, the customer is offered a payment based on the reserve capacity realized or the energy delivered from that capacity. Just like in the case of DLC, the customer receives no prior notice because the emergencies or events are often without warning [162].

A comparison of the various DR programs is listed in Table 2-4.

Table 2-4 Comparison of demand response programs

Incentive-based demand response program			
Components	Direct		Indirect
	Direct Load Control	Interruptible Service	Demand Bidding
Incentives	* Discounted electricity rates * Incentivized payments	Payment based on reserve capacity realized	Revenue from predefined prices and quantity
Initiating party	* System operator * Customers	System operator	Customers
Advanced warning	* No prior notice * Pre-notified (if initiated by customer)	No prior notice	Pre-notified
Non-compliance penalty fees	Yes	Yes	No
Price-based demand response program			
Components	Time-of Use Pricing	Real-Time Pricing	Critical Peak pricing
Tariff category	* Peak * Off-peak	Standard	* Peak * Off-peak
Planning strategy	Set in advance	Real-time	Set in advance
Tariff time-scale	Several hours or days	15 minutes to hourly	No. of peak days in a year
Tariff Type	* Static (fixed) * Dynamic (varies)	Dynamic	* Static * Dynamic
Seasonality	Optional	No	Yes
Drivers	Customers' load pattern	Wholesale market price	* Real-time energy market * Contingency or event

2.6 Supply side management

SSM refers to the efficient generation, transmission, and distribution of electricity to meet customer demand [59]. Efficiency in this context describes supplying reliable and quality power at an optimized price with reduced impact to the environment. SSM must also satisfy demand without unnecessary infrastructure investments [139]. SSM also refers to optimizing supply from various sources to respond to base and peak demands [58].

To match demand with supply, alternative generation must be factored into system planning when the primary source is insufficient. System operators often have primary protocol which involves a response by synchronous generator(s) to a shortage in supply by providing additional power to match consumption. This response is almost always instantaneous. Alternative energy can be supplied from rechargeable batteries, power generators, and other RES sources. Bear in mind that battery storage can be either a SSM or DSM response depending on the state of charge (SoC) of the battery. In other words, If the load requirement is not being met, batteries can act as back-up and supply the load. On the other hand, if excess power is being generated, batteries can behave like a load and absorb power to recharge.

By increasing the energy supply, SSM ensures system reliability by providing power to customers whenever demand increases for example initiating battery storages, ramping up generator production etc. Conversely, supply is reduced to prevent wastage of excess supply generated, and to reduce emissions and operations costs [166].

2.6.1 Back-up power and supply side management: Batteries and generator dispatch

The availability of back-up power is critical to SSM especially when RES are used as the primary generators. To ensure power stability in both off-grid and on-grid PV-connected systems, the hybrid PV system and the battery system are deployed [167].

The hybrid power system utilises electrical energy input into a MG from conventional sources like coal, gas, petrol or diesel. Other energy inputs may include RES and nuclear [70]. Typically, in areas where grid extension is not economically feasible, stand-alone RES and diesel generators have been deployed to meet load demand [82]. Communities served may be rural or insular, and DDGs are commonplace in the local energy system [168].

There are many studies [70,82,169–171] that include generators as part of the MG configurations. In a case study of an RSHMG, Ekpe and Umoh [70] propose including a generator array as a back-up to the PV array and the grid as a measure to improve the grid's reliability due to the erratic nature of power supply from the central grid. This shows that a generator is a viable energy source in maintaining grid reliability. Tsai et al. [170] perform a techno-economic analysis of stand-alone diesel

system, stand-alone PV/storage system, PV/diesel hybrid system (RHMG), PV/diesel/storage hybrid system for the Pratas island in Taiwan. The results of the analysis revealed that the PV/diesel hybrid system configuration had the lowest cost of energy (CoE) at 0.3569 \$/kWh. Murugaperumal et al. [171] validated the design and techno-economic feasibility of an RHMG that uses a diesel generator. The research concludes that the combination of these sources increases system reliability. Pujari and Rudramoorthy [82] proposed using a diesel generator to supply customers under peak load conditions when RES is limited. In an analysis of six different hybrid combinations of PV, battery, wind turbine, and a diesel generator supplying a load demand of 332.97 kWh/d, the RHMG system of PV, diesel generator, battery had the lowest net present cost (NPC) and CoE. Although crucial for system stability and dependability, diesel generators are sometimes viewed and modelled as "a black box" that raises operating costs and has a negative impact on the environment due to their excessive fuel consumption [172].

Battery Energy Storage Systems (BESS) can store energy from a variety of sources and discharge it as needed. Rather than wasting electricity, BESS enables excess generation to be stored when demand is low and used later at a more critical time. The flexibility created from this approach leads to a reduction in cost for the user. In addition, BESS can be charged during off-peak periods to take advantage of reduced utility pricing while supplying its energy as an alternative during peak periods. Both scenarios are widely used by electric vehicles in a vehicle-to-grid model. For example, Zeynali et al. [173] uses a two-stage stochastic programming in a smart home application to lower an average household's cost of purchasing electricity. In this case, a home energy management system coordinates a BESS and the vehicle-to-home capabilities of an electric vehicle. In another example, Song et al. [174] use a multi-objective approach to reduce running costs and improve user comforts by adjusting home energy management scheduling, photovoltaic integration, and battery energy storage integration. By using RTP tariffs in the scheduling, cost reductions of about 40% were realised.

BESS is also used in combination with other sources to reduce the overall CO₂ and electricity costs of the system. For instance, Kusakaka et al. [175] uses a BESS to reduce the fuel consumption of a diesel-powered RTG crane, thereby reducing the operations costs by up to 40%. BESS tend to have higher up-front costs than other

hybrid alternatives and lose the ability to hold a charge as time goes on. This cost and reliability flaw makes BESS the weakest link in the PV system [176–178]. In mature energy price markets with high BESS penetration, the benefits of owning a BESS come from the price difference between times of energy scarcity (or high demand) and times of abundant inexpensive energy [179]. Therefore, the higher the adoption rate, the smaller the price difference which will make buying a BESS less appealing. To mitigate against cost and reliability issues, an EMS is often used to optimally charge and discharge the battery. The EMS will prevent excessive charging, discharging and heat, to prolong the BESS life while charging when electricity rates are low and discharging when they are high.

2.6.2 Flexible power supply generation

Flexibility in power supply is the ability of the system to ramp up or down power generation in response to load demand. It can be done manually or automatically, as part of a strategy or a response to an isolated event. Dispatchable generators, load-following generators, and peaking power plants are all synonymous with flexibility and can shed power generation as part of SSM. Gas turbines, modern coal plants, controllable fuel-based generators, BESS, pumped hydro storage are examples of dispatchable generation units that can be switched on or off depending on load requirements, occurrence of an event, or an economic dispatch strategy [180,181].

While solar and wind are widely viewed as non-dispatchable energy sources because they are not meant to be switched off, they can be curtailed when their penetration levels create an instability in grid. In general, curtailment refers to using less wind or solar energy than is possibly available at a given time [182]. Other causes of curtailment are grid congestion and energy supply overproduction [183]. The key concept is that any power source that can be controlled by increasing or decreasing its output in response to demand at a given time fulfils the criteria for SSM.

2.7 Energy management system

Home Energy Management System (HEMS), Integrated Energy Management System (IEMS), Smart Energy Management System (SEMS) or Centralized Energy Management System (CEMS) are synonymous with EMS and are classified as

systems that optimize SSM and DSM techniques to facilitate the production and use of reliable and cost-effective energy. Historically, EMS referred to management of energy and improvement of energy efficiency. EMS are a set of computer tools that monitor, regulate, and improve the generation, transmission, and distribution of power [69]. The EMS accepts conditions and constraints from either the supply and demand side (or suppliers and consumers) and optimally schedules the consumers load within those constraints [166]. According to Ma et al. [72], the EMS handles monitoring, communication and bi-directional interaction between the source and the load. Historical grids were mainly supplied by single sources with radial distribution configurations. The grid has since evolved with advancements in technology and communication. For smart grid applications, Zhao et al. [49] emphasizes that EMS uses tools like SM, sensors, and other detection devices to achieve five main objectives: minimizing costs, load curve optimization, reduction of CO₂ emissions, maximizing renewable energy outputs, and improved user comfort. These objectives are classified under the three broad frameworks of economy (cost savings), environment, and human comfort.

EMS is broadly classified into two categories: predictive energy management system (PEMS) and real-time energy management system (REMS) [147]. PEMS involves using historical data to generate a load forecast, an energy supply forecast, or a combination of both, to make sure supply optimally matches with demand. However, because forecasting is not a 100% science, real time scheduling of the load needs to be integrated to adjust the prediction errors. As the name suggests, REMS uses real-time algorithms to adjust the control of the load or supply based on the SSM or DSM parameters.

2.7.1 Energy management system control

The EMS can be sometimes called the control system [65]. EMS control is mainly classified into three categories: centralized controllers, decentralized schemes and distributed control strategies, with the centralized controllers more widely proposed for systems that require integrating forecasting with SR and DR [72]. The EMS with a centralized controller works by having a direct connection with each distributed energy resource in the system. The forecasting input into the EMS controller can be in the form of supply forecast, load forecast or a combination of both. It is also able to monitor

and process information, store data and initiate the required SR and/or DR responses based on pre-determined objectives or constraints.

Deka et al. [58] uses a single-phase digital Programmable Intelligent Computer (PIC) microcontroller to regulate the power flow to individual customers. The software was developed in C- language and compiled by MikroC compiler. The controller is programmed with a pre-determined load limit that if exceeded by the customer, results in power being cut-off to that customer. When this is adjusted below the pre-set value, power is restored back to the user. The research points out that the controller was selected because of its “performance, power efficiency and design flexibility” [58].

Shakeri et al. [147] explores a HEMS that uses smart plugs and a local controller. The local controller is the brain of the system. It uses electricity price, customer’s preference and the batteries’ SoC to determine if load should be supplied by the utility, battery or postponed to a later timeslot with lower or off-peak electricity prices. Smart plugs are the interface between the local controller and load and can measure and monitor power consumption of connected electrical appliances. They communicate this information to the local controller and receive feedback on what appropriate control action to implement.

While the above method is effective for power supply that is constant, customers enrolled in a system using this technology will encounter blackouts if they exceed the power threshold. The intermittent nature of solar energy suggests supply will always fluctuate and therefore a more robust approach is required. What this research proposes is for the control system to adjust the supply by calling on standby or alternative generation to meet the required demand. The key difference is that the microcontroller makes decisions based on both the real-time supply and demand, and not just the demand.

The EMS highlighted in [128] optimizes the system based on energy storage in batteries, customer’s load consumption and cost-reflected energy purchases. It has as its input the estimated PV energy forecast, the demand forecast and the maximum contracted power. The aim is to optimally decide next day energy usage by minimizing energy cost and efficiently using the batteries. The EMS algorithm prioritizes consumption of the PV, supply of demand above contracted power, and then charging the battery based on excess PV or electricity purchased at low cost.

2.7.2 Smart meters

Smart meters (SM) are an important element in converting an EMS into a smart EMS, playing an integral part in the communication platform of a smart grid [69]. SM are measurement devices that continuously track and log power consumption in real-time at pre-determined intervals [184]. For the consumer, SM are the most important tools in designing a good DSM program [185]. On a very basic level, this is achieved by providing the customer with their consumption information to influence how they manage it. For the service producer or supplier, the ability to remotely access or control SM reduces operational costs, determines revenue, and improves security in energy supply. In addition, the accuracy of supply and demand forecasts depends on the SM [186].

SM performs other tasks including managing electricity and scheduling appliance usage [187], reducing energy theft, increasing energy efficiency [188], automated data collection [189], and fraud detection [190] to name a few. Smart meters enable a two-way communication between customers and producers and offer flexibility in demand side control. This bi-directional communication is the most important feature of SM [191].

2.8 Integrated energy management system

EMS and IEMS are sometimes used interchangeably [192,193]. Jabbour et al. [194] also switch between the two acronyms but qualify the term “integrated” as an approach that addresses a multi-objective optimisation problem for energy saving, user comfort and maximising energy from renewable energy sources. However, it is important to properly differentiate them based on function and scope. EMS’s definition historically came from the management of a single energy supply source, Wang et al. [195] defines an IEMS as a platform that integrates energy dispatch and control for multiple energy sources. The aim of the IEMS is to achieve reliable and economical operation of integrated energy systems (IES) while meeting the system constraints and realizing optimal scheduling. The authors site the constraints of dealing with multi-energy systems as the reason IEMS, with the development of *Energy Internet* and the use of big data, is a necessity. According to Zhang et al. [196], the enhancement of building performance by existing EMS strategies was done in two silos: building simulation and

control management. This led to longer operating times and inaccuracies. The proposed IEMS corrects this by integrating the physics of the building, renewable energy systems, energy storage, energy distribution systems, heating and cooling technologies, allowing the flexibility of control strategies based on the users' objectives. This suggests that the IEMS can manage multiple energy sources and control strategies. Yi et al. [197] firstly define IES as a system that is made up of numerous subsystems involved in the generation, distribution, and storage of various energy sources. IES integrates multiple energy and power sources using advance energy conversion technologies to efficiently improve energy use, increase system flexibility and reduce carbon emissions [198,199]. Just like EMS, IES is sometimes used interchangeably with IEMS. Reviewing the literature, IES is often defined as an optimization problem with set function objectives. For example, Guo and Xiang [200] propose an interactive Integrated Energy System Planning (IESP) platform that aggregates various energy sources (electricity, gas, heat) and takes as its input investment decisions, load profiles, meteorological data, and techno-economic parameters converting into a mixed integer linear programming problem. The objective is the optimal component size of each energy source, optimised total annual cost and carbon emissions. Huang et al. [201] propose the park integrated energy system (PIES), which is a stochastic optimal scheduling method that aims to achieve low-carbon but economically optimal results while considering generation-side and customer-side uncertainties and their impacts on the system. These uncertainties are comprised of factors such as the variable nature of RES and fluctuations in multi-energy load demands influenced by weather and seasonality, geographic location and so on. Like Huang et al. [201], Feng et al. [202] consider the optimization of the IES considering RES and energy demand uncertainties using the information gap decision theory method. He et al. [203] solve the annual cost of planning for an IES as the objective of its optimization problem by considering electric vehicles swapping stations and carbon capture power systems while using natural gas, heating networks, PV and wind energy on the supply side. The focus of these models is on energy conversion within constraints and not on the IEMS components highlighted in Section 2.2.

Another category of IES to consider are energy system modelling software. Software like PLEXOS use multi-objective decision optimization particularly to solve unit commitment (UC) and energy dispatch (ED) problems. For example,

Papadopoulos [204] highlights a case study of a multi-objective problem maximising the profits of an operator's generation portfolio, minimizing total system costs, and minimizing total generated emissions. These objectives are characterised by Pareto optimality, where the solutions to the objective problem are Pareto solutions in which the improvement of an objective does not diminish another. Here again the objectives and constraints relate to the energy conversion and use. In a review of 75 other energy and modelling software, Ringkjøb et al. [205] classifies them based on technological and economic parameters, general logic, and spatiotemporal resolution. General logic is further broken down into purpose (power system analysis, operation decision such as ED and UC, investment modelling), approach (top-down or bottom-up) and methodology (simulation or optimisation). Most software are black boxes and not openly sourced, with no description of their architecture or methodologies. While some software are classified under supply/demand modelling approaches, this is related to energy markets and not management of supply and demand within the IEMS. Finally, the authors mention that only a few models like EMPS and E2M2 account for uncertainty of VRE sources. The authors highlight “demand side” as an area of future model development suggesting the focus has been more on the supply side. Two of the most applied software in 100% RE systems are LUT Energy System Transition model and EnergyPLAN [206]. These models develop optimised solutions while performing interconnected multi-node, full-hourly, multi-sector energy modelling analysis [207]. LUT Energy System Transition model integrates power, heat, transport, and other industry processes into an IES, using linear optimisation to match total annual energy generation to demand on an hourly basis. While synthesised power load data is one of the model's key inputs, DSM is not considered a key input. EnergyPLAN is a deterministic model that focuses on designing national or regional energy planning strategies. Just like LUT model, EnergyPLAN is ideal for intermittent RES because it produces hourly simulations but does not take DSM inputs, taking actual demand instead. Other more recent energy modelling tools like H2RES use linear optimisation to minimise the discounted yearly operation and system costs while providing long-term capacity investment and dispatching optimisation [208].

Yi et al. [197] describe IEMS as a subcategory of IES, where the “management” term deals with controlling energy distribution under different price conditions. Ren et al. [209] idea of an IEMS is the combination of smart power consumption and IoT into

a single framework. Through IoT, various utility data collection nodes including utility type, consumption cost, predicted consumption patterns and so on are transmitted to their respective databases in the designated centralized data server through a wireless network. This is more efficient than the old EMS method of transmitting smart meter data to centralized storage through GPRS communication. From the above definitions of IEMS, we can broadly infer two things. That IEMS is an integration of multiple energy systems and multiple control strategies leveraging on technology like IoT. Multiple energy systems may refer to varied energy sources like hydro, biomass, solar storage, spinning or stand-by reserve plants and so on while multiple control strategies could involve supply and demand optimisation algorithms, forecasting, and dispatch models. EMS may relate to just one energy system and can be either a PEMS or REMS, while IEMS can implement PEMS through RES forecasting and make adjustments in real-time (REMS). Finally, the concept of integrated demand response (IDR) put forward by Huang et al. [210] is introduced due to changes in technology and the energy market. The authors subscribe that the traditional DR and its single strategy of power system scheduling and control is not sufficient for future grid networks which have developed into multi-energy systems with varied forms of energy consumption, storage, and technologies like combined cooling, heat and power (CCHP). In essence, just like DR has given way to IDR in the context of multi-energy systems and control, IEMS must succeed the EMS in today's complex power networks. On the strength of this, we define IEMS as a system that manages multi-source or multi-energy systems by leveraging on advancement in technology and communication to integrate both PEMS and REMS controls, and initiate supply and demand responses with the aim of balancing the load and power supply in the grid.

Several studies [59,60,166,211–213] have been identified that combine SSM and DSM, or SR and DR, into an EMS to balance supply and demand in a system. In their approach, Deka et al. [211] consider both SSM and DSM strategies supported by an IEMS for data centres. The study chooses to explore the various scenarios of demand side flexibility rather than the actual response. It also explores the operational considerations of consolidation, shifting, migration and frequency scaling in participating in DR programs.

In Karunanithi et al [59], three DSM scenarios were analysed: energy conservation, peak load shifting and a simultaneous combination of both. On the other

hand, the three SSM strategies considered are reduction of T&D losses, increasing the efficiency of all thermal power plants and the simultaneous combination of both.

Monyei and Adewumi [166] use carbon capture and sequestration technology as SSM techniques to reduce operations and emissions cost. On the DR side, RTP and TOU are used to reduce the consumer's electricity bill.

Luo et al. [212] looks at the integration of SSM and DSM in multi-energy systems in buildings. The research focuses on its SSM by using a tri-generation of using the prime mover set of solid oxide fuel cell-gas turbine with cool storage, heat storage and electricity storage managed by an optimization algorithm. Utility electricity was used as the back-up. For DSM, electric cars and other dispatchable appliances were scheduled using predictive modelling with inputs from weather data and other building information. This predictive modelling is part of the demand optimization algorithm. In essence, both supply and demand algorithms would simultaneously predict supply and demand patterns in a day respectively and make sure they coincide.

Ghiasi et al. [213] uses adaptive fuzzy control to regulate supply from the combination of PV, fuel cells, plug-in electric vehicle, the grid, and battery energy storage as the supply sources. DSM strategies use game theory to schedule controllable loads during off-peak periods. Another research that includes the integration of an electric vehicle (EV) was conducted by Aoun et al. [60]. The researchers consider the use of an intelligent EMS, EV integration with smart charging and discharging and load scheduling of appliances as part of the demand response. On the supply side, the integration of a PV system with and without storage was analysed. The sources of energy to the were PV, diesel generator, utility grid and power from the EV batteries.

On another note, there are EMSs that combine energy supply forecasting with either DSM [62,126] or SSM. Very few have investigated all three and their integration into an EMS. In research for the mining industry, Ortiz et al. [214] focuses on using PV and BESS as the primary source of SSM using a mining case study while having the grid as a back-up. DSM technique included managing the hardness of rocks fed to the semiautogenous grinding mill by sending hard rocks (which are harder to grind) into the mill when energy costs are low and sending soft rocks when costs are high. In

addition, solar radiation forecast analysis is done over 13 years of data by combining a trend analysis and a forecast model to determine probabilities for different day types, and hence the marginal probabilities.

Su et al. [129] includes both RE and load forecasting along with an integrated supply-demand side management response. The study uses the Non-dominated Sorting Genetic Algorithm-II (NSGA-II), to solve the multi-objective optimization that includes minimizing energy consumption, maintaining a gas network buffer, improving profits by scheduling loads with best electricity prices, and reducing operational power consumption. The supply side was simulated through an energy network simulation module that included energy sources from PV, wind, the utility grid, and a natural gas network. SSM included an energy conversion model of gas-fired power generation for times of low supply, and a power to gas system for times of excess power production. The DSM framework is developed based on dynamic pricing.

In the research by Pascual et al. [61], an energy management strategy incorporates both battery and thermal storage along with DSM and forecasting techniques. The research analyses a combined thermo-electrical microgrid with wind, PV, solar concentrators and Utility as the primary sources and hot water storage and battery storages as the backup. The SSM relied on the EMS to call up power from battery bank and utility when RE production is low. The DMS's unique strategy is to absorb excess power to the electric heater during peak production, thereby converting the excess power to thermal storage.

Pawar et al. [66] uses an Intelligent Smart Energy Management System (ISEMS) made up of three stages: PV data collection and generation, forecasting model integrated with smart energy management gateway and smart sockets to turn on and off appliances, and an internet-of things (IoT) environment for users to view energy details and manage appliance priorities. The ISEMS manages the system by checking if current demand exceeds a pre-defined maximum demand. If this condition is met, the ISEMS decisive algorithm will turn-off appliances to reduce the demand based on a priority order determined by the user. The ISEMS will also notify the user through the IoT environment when loads are being used during peak periods so that the user can change/update priority requests.

In a study by Shivam et al. [11], a PEMS is deployed using a three-level hierarchical control method which are monitoring and prediction, multi-objective optimization and control of PV-battery hybrid system. A PEMS controller is connected directly to the PV and battery bank, and connected to the grid via a smart meter, drawing from all three sources to satisfy the load based on its algorithm. The PEMS monitors and collects PV, load and SOC data, separating them into day or night by logic control. The output is then fed into the machine learning forecasting framework resulting in PV and load forecasts. Multi-objective optimization is used on the predicted values to find the SOC limit for the battery bank based on maximum battery SOC limit, cost of electricity and maximum allowable CO₂ emissions. The load demand is then satisfied using PV energy, battery bank or the grid. If excess energy is generated that surpasses the battery bank's scheduled maximum SOC limit, then it is sold to the utility grid. Table 2-5 shows a summary of IEMS related studies.

Table 2-5 A summary of studies on integrated energy management systems

Research	Power System	Primary	Backup	EMS - Control System	Prediction Time	Solar Energy Forecasting Methodology	Forecast Outcome	PV Power prediction	SSM	DSM
[126]	74kW DC Microgrid	Wind-PV	Utility	EMS with adaptive nonlinear control theory	24-hour-ahead	Support Vector Machine (SVM)	* Wind Speed * Solar irradiation	* Solar cell power equation * Wind generator power equation		Direct Load Control (DLC) with Sliding Mode Control (SMC)
[56]	* 3.75kW PV * 6.4kWh Battery	PV	Battery	HEMS	24-hour-ahead	N/a	Solar power generated	N/a	Rule-based control strategy to manage energy sharing	Optimisation based strategy: DLC & RTP, DLC & TOU
[62]	Hybrid Microgrid * 25kW PV * 15kW Wind Turbine * 300kW Diesel * 30kW Fuel Cell * 30kW Battery * 30kW Grid	Water, Wind, PV, Fuel cell, Battery, Diesel generator	* Utility * Battery	N/a	24-hour-ahead	* Probability Density Functions (PDF) - Solar * Rayleigh distribution - Wind	* Wind Speed * Solar irradiation	* Rayleigh model for wind power * Solar power output formula	Power Dispatch algorithm with control	Incentive-based DR

	* 30kW Water Micro Turbine									
[65]	Hybrid Microgrid * 2.3kW Wind Turbine * 3.3kW PV * 22kWh Battery	Wind-PV	* Utility * Battery	HEMS algorithm	5-min- ahead	* Wavelet Transform (WT) * Artificial Neural Network	* Wind Speed * Solar irradiation	* Power curve of wind turbine (kw vs m/s ²) * Solar formula to determine I _{sc} and V _{oc} . P= I _{sc} x V _{oc}	Control algorithm: if load > net power (including storage), accept from grid and vice versa	TOU
[128]	640W PV	PV	* Utility * Battery	EMS using 3 inputs.	24-hour- ahead	Random Forest: Non-linear multiparameter regressor	Solar power generated	N/a	Control algorithm: load consumption, if demand > max contracted power, supply from storage	
[215]	6.6 kV distribution network	Grid	PV	HEMS	24-hour- ahead	Just-in-time modelling scheme	Solar power generated	N/a	Controller and voltage regulator	TOU, DLC

[66]	N/a	PV	Utility	Intelligent Smart Energy Management System (ISEMS) with 3 stages	N/a	Particle Swarm Optimization (PSO) based Support Vector Machine (SVM) regression model	Solar irradiation	N/a		TOU for schedulable or chargeable loads
[11]	14.52kW PV	PV	* Utility * Battery	PEMS	1-hour-ahead	*Residual Dilated Causal Convolutional Network (Res-DCCN), *Naïve, *Support Vector Machine (SVM), *Artificial Neural Network (ANN)	Solar power generated	N/a	Control algorithm: if demand > PV, use battery bank. If demand > PV + battery bank, buy from Grid.	Dynamic pricing
[129]	* Wind farm * PV * Natural gas network *Utility	* Wind farm * PV * Natural gas network *Utility	Gas to power	N/a	N/a	Probability Density Function (PDF) based on kernel method using Monte Carlo Simulation	* Wind Speed * Solar irradiation	* Wind power production formula * Real power production of PV and the combined efficiency of PV formulas	The Non-dominated Sorting Genetic Algorithm-II (NSGA-II) with power to gas	Dynamic pricing

[61]	Electro-thermal Microgrid * 6 kW Wind farm * 6 kW PV * Utility * 2 kW solar thermal collectors * 27 kWh Battery * 800L Hot water tank	* Wind farm * PV * Utility * solar thermal collectors	* Utility * Battery * Thermal	EMS using a Central Moving Approach (CMA)	24-hour-ahead	Weather forecasting with NWP model	* Wind Speed * Solar irradiation	* Estimated output power PV equation * Estimated output power Wind equation	Use power from batteries or grid to supplement RE production	Control strategy to increase power to electric heater during peak production, thereby converting excess power to thermal storage
[13]	5 kW Micro Grid	5 kW PV	Pumped hydro storage (PHS)	EMS algorithm	15-min-ahead	Wavelet decomposition-Generalized Neural Network (GNN) based online forecasting model	Solar irradiation	PV power output estimation formula	Control algorithm: if demand > PV, use PHS. If demand > PV + PHS, buy from Grid.	TOU
[147]	* 800W PV * Utility * 4.8 kWh Battery Storage	Utility	* 800W PV * 4.8 kWh Battery	HEMS algorithm	Intra day	N/a	N/a	N/a	Control algorithm: check electricity price. If cheap, operate on utility, if not check battery	TOU

									SOC. If load < SOC, use battery. If not operate on utility or postpone	
[216]	Microgrid * 2.3kW Wind Turbine * 3.3kW PV * Utility * 22kWh battery storage	* 2.3kW Wind Turbine * 3.3kW PV	* Micro gas turbine (MGT) * Electric Vehicle (Mobile Storage) * Battery	HEMS using mixed integer linear programming (MILP)	24-hour- ahead	Enhanced Differential Evolution based Artificial Neural Network (EDE- ANN)	N/a	N/a	N/a	N/a
[217]	* 5 Kw PV * 5 kW Battery	PV	* Utility * Battery	HEMS algorithm	24-hour- ahead (PEMS) 5-min- ahead (REMS)	Robust Self- Attention Multi- Horizon (RSAM)	Solar irradiation	N/a	Schedular algorithm	RTP, TOU, CPP
[218]	*40W PV * 48V 5Ah Li- ion Battery	PV	Battery	Simple Electric Utility Platform (SEUP)	1-hour- ahead	Long Short-Term Memory (LSTM)	Solar power generated	N/a	Use power from batteries to supplement RE production	DLC

[143]	* 1.5MW PV * 0.5MW/1MWh Battery	PV	Battery	SunDial	Multi-hour ahead	N/a	Solar power generated		Use power from batteries to supplement RE production	DR
[67]	* 1 kW PV * 1 kW Wind Turbine * Battery	Wind-PV	Battery	N/a	Ultra-short, short, medium, and long	Random Forest	* Wind Speed * Solar irradiation	*PV power module equation *Wind power module equation	Use power from batteries to supplement RE production	DR

In addition to the power system in which the IEMS is deployed, Table 2-5 also classifies power supply into primary and back-up to highlight what type of system configuration is being used. In column 5, the IEMS control system describes the algorithm by which the IEMS selects both a SSM and DSM response.

The forecasting methodology applies to the algorithm used to arrive at the forecasted outcome. PV power forecasted is broadly divided into three broad categories: direct, indirect and hybrid models [219]. The direct method does not require any internal data from the PV system and determines the PV power directly by relying more on historical data. The indirect model forecasts solar radiation first, and then uses analytical power equations to determine the PV generated based on the power plant's technical parameters. The hybrid model is simply a combination of the direct and indirect models. Finally, the last two columns give the SSM and DSM responses used in the IEMS. Due to the different types and approaches of SEF, SSM and DSM, it is obvious to see that there are many possible combinations. The studies choice of which method to use depends on constraints and objectives unique to each location or case study.

2.9 IEMS perspective and future research

The outcome of Table 2-5 shows that there is potential for different combinations of SSM and DSM, with SEF within an IEMS. Other potential research areas can arise by simply varying any of the parameters in Table 2-5. As stated before, although the choice of SEF, SSM and DSM are determined by constraints and objectives, there is potential to investigate the different combinations and how they affect the power system. For example, does DLC work better by itself than a combination of DLC and TOU? This can lead to other research questions like what are the savings and costs of each type of SSM or DSM, and how does this affect the customer or system operator? A possible research methodology would be to analyse consumption and costs of each SSM or DSM response added to the system against a baseline with no SSM or DSM responses.

IEMS is not limited to PV, and research can be extended to other predictable RES like wind and tidal energy. Other research areas could include integrating artificial intelligence (AI) within an IEMS to support the decision making of a system operator. The ability of AI to sort through different objectives and constraints, while determining the best possible control strategy will be crucial in IEMS adoption across all levels of the grid. From a policy standpoint, it would be important to know the barriers that exist in sharing, accessing, and controlling user load preferences in determining what DSM strategy to implement. With the IEMS's need to access all or most of the system's databases, much thought needs to be given to system security and the protection of both the user and system operator. Finally, in the event of an IEMS fault or collapse, what back-up measures can be set-up to keep the fault from extending to the grid.

To buttress the point of the significance of different DR combinations, Papadimitriou et al. [220] examines DR programs used in several energy hubs studies, and compares them based on factors like cost, dependability, and adaptability. The authors found out that different DR programs were used to achieve different objectives like minimizing costs and/or power interruptions. DR programs have advantages in different settings and iterations must be carried out to determine which DR option gives the best outcome for a particular objective. For instance, while incentive-based DR programs like DLC showed great prospects in reducing power peaks in the energy hubs, they tend to increase the annual cost for the consumer. In terms of performance, TOU pricing is the most popular time-differentiated electrical tariff to reduce load demand and achieve efficient levels of electricity [221–224], it does not always produce the best results. For instance, in a study by Nourollahi et al. [225], RTP reduces the operation cost of a conventional AC microgrid and a hybrid AC-DC microgrid by 8.1% and 53.89% respectively compared to TOU (7.1% and 42%). In another review of optimal charging and scheduling of EVs [226], RTP showed the most promise over TOU, CPP, and peak time rebates. In a study by Cui et al. [227], due to the high billing instability and risk of price fluctuations, DLC in cooperative gaming was favoured over RTP in participating in a balance power market.

Looking at another angle, the absence or presence of PV as either a primary or secondary source affects the way the IEMS operates. In a study by Rastegar et al. [228], PV/utility with TOU improves a base case without DR by 65%, while Utility and TOU without PV improves the base case by 61%.

Different SRs and DRs can be classified either as predictive or real-time responses. For instance, the static TOU is used for longer time frames and is set in advance (predictive) while RTP is determined close to real-time load consumption [229]. By initiating predictive and real-time responses, the IEMS can be both a PEMS and REMS. A valid question emerges as to which of the EMS schemes is better and is it advantageous to have both. To answer that question is to look at both independently. The ability to forecast energy supply or load demand accurately is a crucial part of PEMS, leading to better customer preparation and participation, informed decision-making as well as efficient power system operations and reduced costs [230]. For instance, a forecast of low energy supply gives the PEMS or customer plenty of time to put effective contingency plans in place. But forecasting methods are not 100% accurate due to the stochastic nature of load and PV generation [231]. The result of this is that PV and wind energy forecast errors increase absolute levels of real-time energy imbalance and day-ahead or intra-day spot prices [232]. To mitigate against this, a real-time control component of the EMS must be included to adjust the responses of the load and supply. Therefore, the integration of PEMS and REMS, also referred in some literature as an IEMS, is crucial in maintaining the supply-demand energy balance in the system while minimizing errors. In a study by Hafiz et al. [231], Long Short-term Memory (LSTM) - based deep learning neural network is used to forecast load and PV generation profiles. The deviations from the forecasted profiles are corrected by integrating an offline multistage stochastic optimization model with dual dynamic programming and a real-time based controller.

In another study, Jiang et al. [233] refers to the IEMS as a “double-layered coordinated control approach” and calls PEMS the “schedule layer” while the REMS is the “dispatch layer”. The schedule layer takes as its input forecasted

energy, load data, and market prices and schedules the load sequence. It also creates a power reserve for each time-step to compensate for any forecasting errors. The dispatch layer follows the load sequence set by the schedule layer by optimising power flow and voltage limits, while activating the reserved power to correct the forecasting errors. In another study by Nge et al. [234], the authors use the REMS to compensate for PV forecasting errors by using method of Lagrange multipliers to form an optimal dispatch function to adjust battery power based on market prices. In yet another study, Elkazaz et al. [235] uses a hierarchical two-layer EMS. The first layer is a model predictive control layer which optimizes energy usage by efficiently scheduling home appliances. The second is a real-time controller layer which minimizes energy wastage due to forecast errors by optimally dispatching the BESS. The second layer achieves this by prioritizing self-consumption from the BESS when the load is greater than supply. Conversely, if the supply is greater than the load, the BESS is charged first before excess supply is exported to the grid. In the same study, the findings demonstrate that, for the same battery capacity, the two-layer EMS produced PV self-consumption of up to 91.1% annually, as opposed to a single layer system, which only produced 78.8%. When compared to a single-layer EMS, the two-layer system reduced household payments by up to 27.8% annually. This suggests the two-layer is better than one. With that established, future research lies in investigating the effects in terms of consumption and costs of different SR and DR combinations used in the predictive and real-time layers of the IEMS.

Finally, the methods highlighted in this paper can be compared to other solar integration methods used in applications like UC, ED modelling, and solar energy optimization models. For example, Fang et al. [236] propose a multi-objective UC model that considers the operational risks of load shedding and wind curtailment, to integrate solar energy and optimally allocate power for peak-shaving in a hybrid wind and concentrating solar power plant. Robin and Kory [237] explore the resilience of a microgrid by proposing a stochastic mixed integer model for day-ahead UC that optimises transmission switching, emergency PV generation and power sharing between interconnected grids. Dispatch of multi-energy power systems, like the one proposed by Wang et al. [238], define system

objectives like cost and energy absorption as part of an optimization problem to maximise VREs. A similar optimisation problem can be seen in vehicle-to-grid charging and peer-to-peer energy networks. Grid balancing as described by Aktar et al. [239] is difficult because of the stochastic nature of VRES and load demands. Grid imbalance can be reduced by using electric vehicles as mobile storage units. The authors propose a mixed integer linear optimization algorithm that determines the optimum number and specification of electric vehicles needed to maximise operational and economic benefits. The aim of all these methods is to sustainably increase the use of solar energy in the grid.

2.10 Conclusion

Today's complex power network of multi-energy systems, multi-objectives, diverse load requirements and advancement in technology and communication means that the traditional energy management system (EMS) is not sufficient and must give way to an integrated approach. This paper puts forward the concept of an integrated energy management system (IEMS) as a system that manages multiple energy sources by leveraging on advancement in technology and communication to integrate both predictive and real-time controls, and initiate supply and demand responses to balance the load and power supply in the grid. There is a strong interconnection between solar energy forecasting (SEF), demand side management (DSM) and supply side management (SSM) when deployed in an IEMS according to Table 2-5. From just the simultaneous combination of SSM and DSM, the study by Karunanithi et al. [59] shows up to 18% increase in system reliability. A decentralized solar energy based mini-grid can be a vehicle for solar integration by using an IEMS to match the load to supply. IEMS that manages today's smart grid must be able to interact with both the supply and demand sides of the power system. This paper started with a review of the state of the art of IEMS and following conclusions can be drawn:

- Uncertainty in the power system can be reduced if integrated renewable energy sources (RES) can be predicted. For mini-grids that use solar energy as the primary supply source, it is beneficial to predict the supply

than load consumption since the former dictates how the latter is used. In addition, load consumption is harder to predict due to the unpredictable nature of energy use by consumers.

- Advances in grid technology and communication, a diverse energy mix, and increased interconnections, mean the new EMSs must be able to interact with both the supply and demand sides of the grid. The grid has become smarter so EMSs must also become smarter.
- For efficient energy scheduling and utilization, SEF and an EMS should be integrated with demand side management (DSM).
- Renewable smart hybrid mini-grids (RSHMG) possess the requisite technology and infrastructure for solar energy integration. They also provide viable alternatives for grid-extension.
- The IEMS structure or framework has several possible permutations because of the different types of SEF approaches, SSM and DSM responses. The choice of which method to use is often dictated by constraints and objectives unique to the case study or location.
- Battery energy management system (BESS) is considered an essential part of future hybrid energy systems. However, there are potential issues related to capital and operating expenses, safety risks, and environmental impact, largely due to its capacity size. The use of an IEMS can help minimize these issues by reducing the required capacity, optimizing the charging and discharging processes to maximize efficiency, and preventing overcharging, over-discharging, and overheating of the BESS.
- IEMS can both prevent an energy mismatch between supply and consumption, and in real-time adjust the consumption to the supply in the case of an unforeseen imbalance in the system.

Credit Author Statement

Tolulope O. Falope: Conceptualization, Methodology, Software, Data curation, Formal Analysis, Visualization, Writing – original draft, **Lao Liyun:** Writing – review & editing, Supervision, **Dawid Hanak:** Writing – review & editing, Supervision, **Da Huo:** Writing – review & editing, Supervision.

References

- [1] IEA. SDG7: Data and Projections 2021. <https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity> (accessed August 3, 2021).
- [2] Bollen MHJ, Moreno-Munoz A. The Power Grid as Part of a 100% Renewable Energy System. In: Moreno-Munoz A, editor. Large Scale Grid Integration of Renewable Energy Sources, Institution of Engineering and Technology; 2017, p. 1–27.
- [3] Mahmood D, Javaid N, Ahmed G, Khan S, Monteiro V. A review on optimization strategies integrating renewable energy sources focusing uncertainty factor – Paving path to eco-friendly smart cities. Sustainable Computing: Informatics and Systems 2021;30:100559. <https://doi.org/10.1016/J.SUSCOM.2021.100559>.
- [4] Rehmani MH, Reisslein M, Rachedi A, Erol-Kantarci M, Radenkovic M. Integrating Renewable Energy Resources into the Smart Grid: Recent Developments in Information and Communication Technologies. IEEE Trans Industr Inform 2018;14:2814–25. <https://doi.org/10.1109/TII.2018.2819169>.
- [5] bp. Statistical Review of World Energy 2022. 2022.
- [6] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renewable and Sustainable Energy Reviews 2018;81:912–28. <https://doi.org/10.1016/J.RSER.2017.08.017>.

- [7] Yao X, Yi B, Yu Y, Fan Y, Zhu L. Economic analysis of grid integration of variable solar and wind power with conventional power system. *Appl Energy* 2020;264:114706. <https://doi.org/10.1016/J.APENERGY.2020.114706>.
- [8] Karimi E, Kazerani M. Impact of demand response management on improving social welfare of remote communities through integrating renewable energy resources. *Canadian Conference on Electrical and Computer Engineering* 2017. <https://doi.org/10.1109/CCECE.2017.7946739>.
- [9] Katz J, Cochran J. Integrating Variable Renewable Energy into the Grid: Key Issues Greening the Grid. *National Renewable Energy Laboratory* 2015;NREL/FS-6A20-63033.
- [10] Deshmukh MKG, Sameeroddin M, Abdul D, Abdul Sattar M. Renewable energy in the 21st century: A review. *Mater Today Proc* 2021. <https://doi.org/10.1016/J.MATPR.2021.05.501>.
- [11] Shivam K, Tzou JC, Wu SC. A multi-objective predictive energy management strategy for residential grid-connected PV-battery hybrid systems based on machine learning technique. *Energy Convers Manag* 2021;237:114103. <https://doi.org/10.1016/J.ENCONMAN.2021.114103>.
- [12] International Renewable Energy Agency. *Renewable Power Generation Costs in 2020*. Abu Dhabi: 2021.
- [13] Chaudhary P, Rizwan M. Energy management supporting high penetration of solar photovoltaic generation for smart grid using solar forecasts and pumped hydro storage system. *Renew Energy* 2018;118:928–46. <https://doi.org/10.1016/J.RENENE.2017.10.113>.
- [14] Brandt B, Kletzer E, Pilz H, Hadzhiyska D, Seizov P. *Silicon-Chemistry Carbon Balance An assessment of Greenhouse Gas Emissions and Reductions*. n.d.
- [15] Sun H, Guo Q, Zhang B, Wu W, Wang B, Shen X, et al. *Integrated Energy Management System: Concept, Design, and Demonstration in China*. IEEE

[16] Tiwari A, Pindoriya NM. Automated Demand Response in Smart Distribution Grid: A Review on Metering Infrastructure, Communication Technology and Optimization Models. *Electric Power Systems Research* 2022;206:107835. <https://doi.org/10.1016/J.EPSR.2022.107835>.

[17] Jafari M, Botterud A, Sakti A. Decarbonizing power systems: A critical review of the role of energy storage. *Renewable and Sustainable Energy Reviews* 2022;158. <https://doi.org/10.1016/J.RSER.2022.112077>.

[18] Farag MM, Bansal RC. Solar energy development in the GCC region—a review on recent progress and opportunities. *International Journal of Modelling and Simulation* 2022. <https://doi.org/10.1080/02286203.2022.2105785>.

[19] Formolli M, Croce S, Vettorato D, Paparella R, Scognamiglio A, Mainini AG, et al. Solar Energy in Urban Planning: Lesson Learned and Recommendations from Six Italian Case Studies. *Applied Sciences (Switzerland)* 2022;12. <https://doi.org/10.3390/APP12062950>.

[20] Peng J, Yan J, Zhai Z, Markides CN, Lee ES, Eicker U, et al. Solar energy integration in buildings. *Appl Energy* 2020;264. <https://doi.org/10.1016/J.APENERGY.2020.114740>.

[21] Rubio-Aliaga, García-Cascales MS, Sánchez-Lozano JM, Molina-García A. Multidimensional analysis of groundwater pumping for irrigation purposes: Economic, energy and environmental characterization for PV power plant integration. *Renew Energy* 2019;138:174–86. <https://doi.org/10.1016/J.RENENE.2019.01.077>.

[22] Ozoegwu CG, Mgbemene CA, Ozor PA. The status of solar energy integration and policy in Nigeria. *Renewable and Sustainable Energy Reviews* 2017;70:457–71. <https://doi.org/10.1016/J.RSER.2016.11.224>.

[23] Ozoegwu CG, Akpan PU. A review and appraisal of Nigeria's solar energy policy objectives and strategies against the backdrop of the renewable energy

policy of the Economic Community of West African States. *Renewable and Sustainable Energy Reviews* 2021;143. <https://doi.org/10.1016/J.RSER.2021.110887>.

[24] Khan AA, Minai AF, Pachauri RK, Malik H. Optimal Sizing, Control, and Management Strategies for Hybrid Renewable Energy Systems: A Comprehensive Review. *Energies* 2022;15:6249. <https://doi.org/10.3390/EN15176249>.

[25] Panda S, Mohanty S, Rout PK, Sahu BK, Parida SM, Kotb H, et al. An Insight into the Integration of Distributed Energy Resources and Energy Storage Systems with Smart Distribution Networks Using Demand-Side Management. *Applied Sciences (Switzerland)* 2022;12. <https://doi.org/10.3390/APP12178914>.

[26] Rayid M, Mojumder H, Hasanuzzaman · M, Erdem Cuce ·. Prospects and challenges of renewable energy-based microgrid system in Bangladesh: a comprehensive review. *Clean Technol Environ Policy* 2022;24:1987–2009. <https://doi.org/10.1007/s10098-022-02301-5>.

[27] Apeh OO, Meyer EL, Overen OK. Contributions of Solar Photovoltaic Systems to Environmental and Socioeconomic Aspects of National Development - A Review. *Energies (Basel)* 2022;15:5963. <https://doi.org/10.3390/EN15165963>.

[28] Khan T, Yu M, Waseem M. Review on recent optimization strategies for hybrid renewable energy system with hydrogen technologies: State of the art, trends and future directions. *Int J Hydrogen Energy* 2022;47:25155–201. <https://doi.org/10.1016/J.IJHYDENE.2022.05.263>.

[29] Alami AH, Rabaia MKH, Sayed ET, Ramadan M, Abdelkareem MA, Alasad S, et al. Management of potential challenges of PV technology proliferation. *Sustainable Energy Technologies and Assessments* 2022;51. <https://doi.org/10.1016/J.SETA.2021.101942>.

[30] Lamnatou C, Chemisana D, Cristofari C. Smart grids and smart technologies in relation to photovoltaics, storage systems, buildings and the

environment. Renew Energy 2022;185:1376–91.
<https://doi.org/10.1016/J.RENENE.2021.11.019>.

[31] Khosrojerdi F, Akhigbe O, Gagnon S, Ramirez A, Richards G. Integrating artificial intelligence and analytics in smart grids: a systematic literature review. International Journal of Energy Sector Management 2022;16:318–38.
<https://doi.org/10.1108/IJESM-06-2020-0011>.

[32] Banu A, Bicer Y. Review on CO_x-free hydrogen from methane cracking: Catalysts, solar energy integration and applications. Energy Conversion and Management: X 2021;12. <https://doi.org/10.1016/J.ECMX.2021.100117>.

[33] Klaas L, Guban D, Roeb M, Sattler C. Recent progress towards solar energy integration into low-pressure green ammonia production technologies. Int J Hydrogen Energy 2021;46:25121–36.
<https://doi.org/10.1016/J.IJHYDENE.2021.05.063>.

[34] Stevovic I, Mirjanic D, Petrovic N. Integration of solar energy by nature-inspired optimization in the context of circular economy. Energy 2021;235. <https://doi.org/10.1016/J.ENERGY.2021.121297>.

[35] Dashtpeyma M, Ghodsi R. Developing the resilient solar energy management system: a hybrid qualitative-quantitative approach. International Journal of Ambient Energy 2021;42:1892–911.
<https://doi.org/10.1080/01430750.2019.1630301>.

[36] Shafiul Alam M, Al-Ismaail FS, Salem A, Abido MA. High-level penetration of renewable energy sources into grid utility: Challenges and solutions. IEEE Access 2020;8:190277–99. <https://doi.org/10.1109/ACCESS.2020.3031481>.

[37] Østergaard PA, Duic N, Noorollahi Y, Mikulcic H, Kalogirou S. Sustainable development using renewable energy technology. Renew Energy 2020;146:2430–7. <https://doi.org/10.1016/J.RENENE.2019.08.094>.

[38] Syafaruddin, Zinger DS. A review of hybrid power generation: Modelling-simulation, control strategy and future trend development. Journal of Engineering

Science and Technology Review 2020;13:249–63.
<https://doi.org/10.25103/JESTR.134.26>.

[39] Nižetić S, Djilali N, Papadopoulos A, Rodrigues JJPC. Smart technologies for promotion of energy efficiency, utilization of sustainable resources and waste management. *J Clean Prod* 2019;231:565–91.
<https://doi.org/10.1016/J.JCLEPRO.2019.04.397>.

[40] Azuatalam D, Paridari K, Ma Y, Förstl M, Chapman AC, Verbič G. Energy management of small-scale PV-battery systems: A systematic review considering practical implementation, computational requirements, quality of input data and battery degradation. *Renewable and Sustainable Energy Reviews* 2019;112:555–70. <https://doi.org/10.1016/J.RSER.2019.06.007>.

[41] Guo Z, Sun Y, Pan SY, Chiang PC. Integration of green energy and advanced energy-efficient technologies for municipal wastewater treatment plants. *Int J Environ Res Public Health* 2019;16.
<https://doi.org/10.3390/IJERPH16071282>.

[42] Lauka D, Haine K, Gusca J, Blumberga D. Solar energy integration in future urban plans of the South and Nordic cities. *Energy Procedia* 2018;152:1127–32. <https://doi.org/10.1016/J.EGYPRO.2018.09.137>.

[43] Khoshkbar-Sadigh A, Smedley KM. The necessity of time-series simulation for investigation of large-scale solar energy penetration. 2015 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2015 2015. <https://doi.org/10.1109/ISGT.2015.7131803>.

[44] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. *Prog Energy Combust Sci* 2013;39:535–76.
<https://doi.org/10.1016/J.PECS.2013.06.002>.

[45] David TM, Silva Rocha Rizol PM, Guerreiro Machado MA, Buccieri GP. Future research tendencies for solar energy management using a bibliometric analysis, 2000–2019. *Heliyon* 2020;6:e04452.
<https://doi.org/10.1016/J.HELİYON.2020.E04452>.

- [46] Basmadjian R, Niedermeier F, de Meer H. Demand-Side Flexibility and Supply-Side Management: The Use Case of Data Centers and Energy Utilities 2017;187–204. https://doi.org/10.1007/978-3-319-65082-1_9.
- [47] Cox S, Xu K. Integration of Large-Scale Renewable Energy in the Bulk Power System: Good Practices from International Experiences. National Renewable Energy Laboratory 2020.
- [48] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustain Cities Soc* 2020;55. <https://doi.org/10.1016/J.SCS.2020.102052>.
- [49] Zhao X, Gao W, Qian F, Ge J. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system. *Energy* 2021;229:120538. <https://doi.org/10.1016/J.ENERGY.2021.120538>.
- [50] Kaur A, Nonnenmacher L, Coimbra CFM. Net load forecasting for high renewable energy penetration grids. *Energy* 2016;114:1073–84. <https://doi.org/10.1016/J.ENERGY.2016.08.067>.
- [51] Cai H, Shen S, Lin Q, Li X, Xiao H. Predicting the Energy Consumption of Residential Buildings for Regional Electricity Supply-Side and Demand-Side Management. *IEEE Access* 2019;7:30386–97. <https://doi.org/10.1109/ACCESS.2019.2901257>.
- [52] Marwan M, Naim K, Thaha S, Sofyan S, Idris AR, Asri A. Integrating on-site renewable energy to innovative EDRP in reducing electricity peak demand. 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST) 2021;113–7. <https://doi.org/10.1109/ICAICST53116.2021.9497827>.
- [53] Wang Y, Zhang N, Chen X. A Short-Term Residential Load Forecasting Model Based on LSTM Recurrent Neural Network Considering Weather Features. *Energies* 2021, Vol 14, Page 2737 2021;14:2737. <https://doi.org/10.3390/EN14102737>.

- [54] Sha H, Moujahed M, Qi D. Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings. *Energy Build* 2021;242:110980. <https://doi.org/10.1016/J.ENBUILD.2021.110980>.
- [55] Narvaez G, Giraldo LF, Bressan M, Pantoja A. Machine learning for site-adaptation and solar radiation forecasting. *Renew Energy* 2021;167:333–42. <https://doi.org/10.1016/J.RENENE.2020.11.089>.
- [56] Rochd A, Benazzouz A, Ait Abdelmoula I, Raihani A, Ghennioui A, Naimi Z, et al. Design and implementation of an AI-based & IoT-enabled Home Energy Management System: A case study in Benguerir — Morocco. *Energy Reports* 2021;7:699–719. <https://doi.org/10.1016/J.EGYR.2021.07.084>.
- [57] Aluko OE, Onibonoje MO, Dada JO. A Review of the Control System Roles in Integrating Renewable Energy into the National Grid. 2020 IEEE PES/IAS PowerAfrica, *PowerAfrica* 2020 2020. <https://doi.org/10.1109/POWERAFRICA49420.2020.9219971>.
- [58] Deka K, Mahapatra S, Bordoloi PK, Medhi B. Design and testing of a control system for supply side electrical load management. *Proceedings of 2014 1st International Conference on Non Conventional Energy: Search for Clean and Safe Energy, ICONCE* 2014 2014:277–81. <https://doi.org/10.1109/ICONCE.2014.6808726>.
- [59] Karunanithi K, Saravanan S, Prabakar BR, Kannan S, Thangaraj C. Integration of Demand and Supply Side Management strategies in Generation Expansion Planning. *Renewable and Sustainable Energy Reviews* 2017;73:966–82. <https://doi.org/10.1016/J.RSER.2017.01.017>.
- [60] Aoun A, Ibrahim H, Ghandour M, Ilinca A. Supply Side Management vs. Demand Side Management of a Residential Microgrid Equipped with an Electric Vehicle in a Dual Tariff Scheme. *Energies* 2019, Vol 12, Page 4351 2019;12:4351. <https://doi.org/10.3390/EN12224351>.

- [61] Pascual J, Arcos-Aviles D, Ursúa A, Sanchis P, Marroyo L. Energy management for an electro-thermal renewable-based residential microgrid with energy balance forecasting and demand side management. *Appl Energy* 2021;295:117062. <https://doi.org/10.1016/J.APENERGY.2021.117062>.
- [62] Aghajani GR, Shayanfar HA, Shayeghi H. Demand side management in a smart micro-grid in the presence of renewable generation and demand response. *Energy* 2017;126:622–37. <https://doi.org/10.1016/J.ENERGY.2017.03.051>.
- [63] Kakran S, Chanana S. Smart operations of smart grids integrated with distributed generation: A review. *Renewable and Sustainable Energy Reviews* 2018;81:524–35. <https://doi.org/10.1016/J.RSER.2017.07.045>.
- [64] Guelpa E, Verda V. Demand response and other demand side management techniques for district heating: A review. *Energy* 2021;219:119440. <https://doi.org/10.1016/J.ENERGY.2020.119440>.
- [65] Elma O, Taşcıkaraoğlu A, Tahir İnce A, Selamoğulları US. Implementation of a dynamic energy management system using real time pricing and local renewable energy generation forecasts. *Energy* 2017;134:206–20. <https://doi.org/10.1016/J.ENERGY.2017.06.011>.
- [66] Pawar P, TarunKumar M, Vittal K. P. An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement* 2020;152:107187. <https://doi.org/10.1016/J.MEASUREMENT.2019.107187>.
- [67] Kiptoo MK, Adewuyi OB, Lotfy ME, Ibrahimi AM, Senjyu T. Harnessing demand-side management benefit towards achieving a 100% renewable energy microgrid. *Energy Reports* 2020;6:680–5. <https://doi.org/10.1016/J.EGYR.2019.11.137>.
- [68] Khezri R, Mahmoudi A, Haque MH, Khalilpour K. Energy Management and Optimal Planning of a Residential Microgrid with Time-of-Use Electricity Tariffs 2021:841–6. <https://doi.org/10.1109/ECCE47101.2021.9595208>.

- [69] Saleem MU, Usman MR, Usman MA, Politis C. Design, Deployment and Performance Evaluation of an IoT Based Smart Energy Management System for Demand Side Management in Smart Grid. *IEEE Access* 2022;10:15261–78. <https://doi.org/10.1109/ACCESS.2022.3147484>.
- [70] Ekpe UM, Umoh VB. From solar home systems to smart hybrid mini-grid systems - A nigerian case study. 2021 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2021 2021. <https://doi.org/10.1109/ISGT49243.2021.9372249>.
- [71] Sigarchian SG, Paleta R, Malmquist A, Pina A. Feasibility study of using a biogas engine as backup in a decentralized hybrid (PV/wind/battery) power generation system – Case study Kenya. *Energy* 2015;90:1830–41. <https://doi.org/10.1016/J.ENERGY.2015.07.008>.
- [72] Ma J, Ma X. A review of forecasting algorithms and energy management strategies for microgrids. *Systems Science & Control Engineering: An Open Access Journal* 2018;6:237–48. <https://doi.org/10.1080/21642583.2018.1480979>.
- [73] Perera ATD, Wickramasinghe PU, Scartezzini JL, Nik VM. Integrating renewable energy technologies into distributed energy systems maintaining system flexibility. *Proceedings of the 2018 5th International Symposium on Environment-Friendly Energies and Applications, EFEA 2018* 2019. <https://doi.org/10.1109/EFEA.2018.8617046>.
- [74] Koolen D, Bunn D, Ketter W. Renewable energy technologies and electricity forward market risks. *Energy Journal* 2021;42:21–45. <https://doi.org/10.5547/01956574.42.4.DKOO>.
- [75] NERC. Apply for Mini Grid License 2021. <https://nerc.gov.ng/index.php/home/operators/mini-grid> (accessed August 3, 2021).
- [76] Mosbah F., Iqbal T. Design and Analysis of A community Mini-grid Power System for Libya. *ResearchGate* 2020.

https://www.researchgate.net/publication/347463456_Design_and_Analysis_of_A_community_Mini-grid_Power_System_for_Libya (accessed August 3, 2021).

[77] IRENA. Innovation Outlook: Renewable Mini-grids. International Renewable Energy Agency 2016:1–184.

[78] Zhang J, Huang L, Shu J, Wang H, Ding J. Energy management of PV-diesel-battery hybrid power system for island stand-alone micro-grid. Energy Procedia 2017;105:2201–6. <https://doi.org/10.1016/j.egypro.2017.03.622>.

[79] Robert FC, Sisodia GS, Gopalan S. A critical review on the utilization of storage and demand response for the implementation of renewable energy microgrids. Sustain Cities Soc 2018;40:735–45. <https://doi.org/10.1016/J.SCS.2018.04.008>.

[80] Ravanbach B, Kuhnelt M, Hanke B, von Maydell K, van Dyk EE, Vumbugwa M, et al. Development of a Smart Monitoring and Evaluation Framework for Hybrid Renewable Mini-grids. 2020 15th International Conference on Ecological Vehicles and Renewable Energies, EVER 2020 2020. <https://doi.org/10.1109/EVER48776.2020.9243124>.

[81] Perera ATD, Nik VM, Mauree D, Scartezzini JL. Electrical hubs: An effective way to integrate non-dispatchable renewable energy sources with minimum impact to the grid. Appl Energy 2017;190:232–48. <https://doi.org/10.1016/J.APENERGY.2016.12.127>.

[82] Pujari HK, Rudramoorthy M. Optimal design and techno-economic analysis of a hybrid grid-independent renewable energy system for a rural community. International Transactions on Electrical Energy Systems 2021;31:e13007. <https://doi.org/10.1002/2050-7038.13007>.

[83] Come Zebra EI, van der Windt HJ, Nhumaio G, Faaij APC. A review of hybrid renewable energy systems in mini-grids for off-grid electrification in developing countries. Renewable and Sustainable Energy Reviews 2021;144:111036. <https://doi.org/10.1016/J.RSER.2021.111036>.

- [84] Oladigbolu JO, Ramli MAM, Al-Turki YA. Feasibility study and comparative analysis of hybrid renewable power system for off-grid rural electrification in a typical remote village located in Nigeria. *IEEE Access* 2020;8:171643–63. <https://doi.org/10.1109/ACCESS.2020.3024676>.
- [85] Falk J, Angelmahr M, Schade W, Schenk-Mathes H. Socio-economic impacts and challenges associated with the electrification of a remote area in rural Tanzania through a mini-grid system. *Energ Ecol Environ* 2021. <https://doi.org/10.1007/s40974-021-00216-3>.
- [86] Szewczuk S. Hybrid mini-grids: Smart sustainable energy for low income communities. *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE 2017*. <https://doi.org/10.23919/ICUE.2017.8068012>.
- [87] Archana, Shankar R, Singh S. Development of smart grid for the power sector in India. *Cleaner Energy Systems* 2022;2:100011. <https://doi.org/10.1016/J.CLES.2022.100011>.
- [88] Raza MA, Aman MM, Abro AG, Tunio MA, Khatri KL, Shahid M. Challenges and potentials of implementing a smart grid for Pakistan's electric network. *Energy Strategy Reviews* 2022;43:100941. <https://doi.org/10.1016/J.ESR.2022.100941>.
- [89] Moreno-Garcia IM, Real-Calvo R, Pallares-Lopez V, Gonzalez-Redondo MJ, Santiago I. Towards the Extensive Use of Renewable Energy Resources: Needs, Conditions and Enabling Technologies. In: Moreno-Munoz A, editor. *Large Scale Grid Integration of Renewable Energy Sources*, Institution of Engineering and Technology; 2017, p. 171–210.
- [90] Judge MA, Manzoor A, Maple C, Rodrigues JJPC, Islam S ul. Price-based demand response for household load management with interval uncertainty. *Energy Reports* 2021. <https://doi.org/10.1016/J.EGYR.2021.02.064>.
- [91] Palahalli H, Ragaini E, Gruosso G. Smart Grid Simulation including Communication Network: A Hardware in the Loop Approach. *IEEE Access* 2019;7:90171–9. <https://doi.org/10.1109/ACCESS.2019.2927821>.

- [92] Hall NL, Jeanneret TD, Rai A. Cost-reflective electricity pricing: Consumer preferences and perceptions. *Energy Policy* 2016;95:62–72. <https://doi.org/10.1016/J.ENPOL.2016.04.042>.
- [93] Jordehi AR. Optimisation of demand response in electric power systems, a review. *Renewable and Sustainable Energy Reviews* 2019;103:308–19. <https://doi.org/10.1016/J.RSER.2018.12.054>.
- [94] Álvarez JM, Laita I, Palomo LM, Pigueiras EL, Solano MG. Grid Integration of Large-Scale PV Plants: Dealing with Power Fluctuations. In: Moreno-Munoz A, editor. *Large Scale Grid Integration of Renewable Energy Sources*, Institution of Engineering and Technology; 2017, p. 131–70.
- [95] Kromer M, Zeifman M, Roth K, Yip T. The SunDial Framework: Enabling High Penetration Solar through the Integration of Energy Storage, Demand Management, and Forecasting. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion, WCPEC 2018 - A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU PVSEC 2018:1436–9. <https://doi.org/10.1109/PVSC.2018.8547884>.
- [96] Li B, Zhang J. A review on the integration of probabilistic solar forecasting in power systems. *Solar Energy* 2020;210:68–86. <https://doi.org/10.1016/J.SOLENER.2020.07.066>.
- [97] Falope TO, Lao L, Hanak D. A Unique Three-Step Weather Data Approach in Solar Energy Prediction Using Machine Learning. *Sustainable Energy Solutions for a Post-COVID Recovery towards a Better Future: Part VII*, Bangkok: International Conference on Applied Energy ; 2021, p. 1–7. <https://doi.org/https://doi.org/10.46855/energy-proceedings-9863>.
- [98] Sarp S, Kuzlu M, Cali U, Guler O. An Interpretable Solar Photovoltaic Power Generation Forecasting Approach Using An Explainable Artificial Intelligence Tool. *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2021, p. 1–5. <https://doi.org/10.1109/ISGT49243.2021.9372263>.

- [99] Aprillia H, Yang H-T, Huang C-M. Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network-Salp Swarm Algorithm. *Energies (Basel)* 2020;13:1–20. <https://doi.org/10.3390/en13081879>.
- [100] Cannizzaro D, Aliberti A, Bottaccioli L, Macii E, Acquaviva A, Patti E. Solar radiation forecasting based on convolutional neural network and ensemble learning. *Expert Syst Appl* 2021;181. <https://doi.org/10.1016/j.eswa.2021.115167>.
- [101] Gu B, Shen H, Lei X, Hu H, Liu X. Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method. *Appl Energy* 2021;299. <https://doi.org/10.1016/j.apenergy.2021.117291>.
- [102] Nguyen NQ, Bui LD, Doan BV, Sanseverino ER, Cara DD, Nguyen QD. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. *Electric Power Systems Research* 2021;199. <https://doi.org/10.1016/j.epsr.2021.107427>.
- [103] Pan C, Tan J, Feng D. Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation. *Neurocomputing* 2021;453:552–62. <https://doi.org/10.1016/j.neucom.2020.10.027>.
- [104] Ngoc-Lan Huynh A, Deo RC, Ali M, Abdulla S, Raj N. Novel short-term solar radiation hybrid model: Long short-term memory network integrated with robust local mean decomposition. *Appl Energy* 2021;298. <https://doi.org/10.1016/j.apenergy.2021.117193>.
- [105] Pedregal DJ, Trapero JR. Adjusted combination of moving averages: A forecasting system for medium-term solar irradiance. *Appl Energy* 2021;298. <https://doi.org/10.1016/j.apenergy.2021.117155>.
- [106] Ahmad T, Zhang D, Huang C. Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications. *Energy* 2021;231. <https://doi.org/10.1016/j.energy.2021.120911>.

- [107] Korkmaz D. SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting. *Appl Energy* 2021;300. <https://doi.org/10.1016/j.apenergy.2021.117410>.
- [108] Li LL, Wen SY, Tseng ML, Wang CS. Renewable energy prediction: A novel short-term prediction model of photovoltaic output power. *J Clean Prod* 2019;228:359–75. <https://doi.org/10.1016/J.JCLEPRO.2019.04.331>.
- [109] Najibi F, Apostolopoulou D, Alonso E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *International Journal of Electrical Power and Energy Systems* 2021;130. <https://doi.org/10.1016/j.ijepes.2021.106916>.
- [110] Rodríguez F, Fleetwood A, Galarza A, Fontán L. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renew Energy* 2018;126:855–64. <https://doi.org/10.1016/J.RENENE.2018.03.070>.
- [111] Thukral MK. Solar Power Output Prediction Using Multilayered Feedforward Neural Network: A Case Study of Jaipur. 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2020, p. 1–6. <https://doi.org/10.1109/iSSSC50941.2020.9358821>.
- [112] Wang J, Li P, Ran R, Che Y, Zhou Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Applied Sciences* 2018;8:1–14. <https://doi.org/10.3390/app8050689>.
- [113] Yakoubi H, el Mghouchi Y, Abdou N, Hajou A, Khellouki A. Correlating clearness index with cloud cover and other meteorological parameters for forecasting the global solar radiation over Morocco. *Optik (Stuttg)* 2021;242. <https://doi.org/10.1016/j.ijleo.2021.167145>.
- [114] Zang H, Cheng L, Ding T, Cheung KW, Liang Z, Wei Z, et al. Hybrid method for short-term photovoltaic power forecasting based on deep

convolutional neural network. IET Gener Transm Distrib 2018;12:4557–67. <https://doi.org/10.1049/iet-gtd.2018.5847>.

[115] Chen B, Lin P, Lai Y, Cheng S, Chen Z, Wu L. Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets. Electronics (Basel) 2020;9. <https://doi.org/10.3390/electronics9020289>.

[116] Anaadumba R, Liu Q, Marah BD, Nakoty FM, Liu X, Zhang Y. A renewable energy forecasting and control approach to secured edge-level efficiency in a distributed micro-grid. Cybersecurity 2021;4. <https://doi.org/10.1186/s42400-020-00065-3>.

[117] Kushwaha V, Pindoriya NM. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast | Elsevier Enhanced Reader. Elsevier Ltd 2019;140:124–39. <https://doi.org/10.1016/j.renene.2019.03.020>.

[118] Iyengar S, Sharma N, Irwin D, Shenoy P, Ramamritham K. SolarCast-A Cloud-based Black Box Solar Predictor for Smart Homes. BuildSys '14: 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, 2014, p. 40–9. <https://doi.org/10.1145/2674061.2674071>.

[119] Andrade JR, Bessa RJ. Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. IEEE Trans Sustain Energy 2017;8th:1571–80. <https://doi.org/10.1109/TSTE.2017.2694340>.

[120] Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Math Comput Simul 2017;131:88–100. <https://doi.org/10.1016/J.MATCOM.2015.05.010>.

[121] Persson C, Bacher P, Shiga T, Madsen H. Multi-site solar power forecasting using gradient boosted regression trees. Solar Energy 2017;150:423–36. <https://doi.org/10.1016/J.SOLENER.2017.04.066>.

- [122] Abedinia O, Raisz D, Amjady N. IET Renewable Power Generation Effective prediction model for Hungarian small-scale solar power output. The Institution of Engineering and Technology 2017;11:1648–58. <https://doi.org/10.1049/iet-rpg.2017.0165>.
- [123] Carrera B, Sim M-K, Jung J-Y. PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data; PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data. The Institution of Engineering and Technology 2020;14:2192–201. <https://doi.org/10.1049/iet-rpg.2018.6174>.
- [124] Kim S-G, Jung J-Y, Sim MK. A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning. Sustainability 2019. <https://doi.org/10.3390/su11051501>.
- [125] Kyliashkina IA, Eroshenko SA, Shelyug S. Intelligent Systems as a Tool for Predicting Electrical Energy and Power Generation. 2019 60th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), 2019, p. 1–5. <https://doi.org/10.1109/ITMS47855.2019.8940721>.
- [126] Minhas DM, Khalid RR, Frey G. Load control for supply-demand balancing under Renewable Energy forecasting. 2017 IEEE 2nd International Conference on Direct Current Microgrids, ICDCM 2017 2017:365–70. <https://doi.org/10.1109/ICDCM.2017.8001071>.
- [127] Ronay K, Bica D, Munteanu C. Micro-grid Development Using Artificial Neural Network for Renewable Energy Forecast and System Control. Procedia Eng 2017;181:818–23. <https://doi.org/10.1016/J.PROENG.2017.02.472>.
- [128] Manjarres D, Alonso R, Gil-Lopez S, Landa-Torres I. Solar Energy Forecasting and Optimization System for Efficient Renewable Energy Integration. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2017;10691 LNAI:1–12. https://doi.org/10.1007/978-3-319-71643-5_1.

- [129] Su H, Chi L, Zio E, Li Z, Fan L, Yang Z, et al. An integrated, systematic data-driven supply-demand side management method for smart integrated energy systems. *Energy* 2021;235. <https://doi.org/10.1016/J.ENERGY.2021.121416>.
- [130] Sangrody H, Zhou N, Zhang Z. Similarity-Based Models for Day-Ahead Solar PV Generation Forecasting. *IEEE Access* 2020;8:104469–78. <https://doi.org/10.1109/ACCESS.2020.2999903>.
- [131] Pawar P, Mithulanathan N, Raza MQ. Solar PV Power Forecasting Using Modified SVR with Gauss-Newton Method. *Proceedings - 2020 IEEE 2nd Global Power, Energy and Communication Conference, GPECOM 2020* 2020:226–31. <https://doi.org/10.1109/GPECOM49333.2020.9247935>.
- [132] Kazem HA, Yousif JH, Chaichan MT, Al-Waeli AHA, Sopian K. Long-term power forecasting using FRNN and PCA models for calculating output parameters in solar photovoltaic generation. *Heliyon* 2022;8:e08803. <https://doi.org/10.1016/J.HELİYON.2022.E08803>.
- [133] Munawar U, Wang Z. A Framework of Using Machine Learning Approaches for Short-Term Solar Power Forecasting. *Journal of Electrical Engineering & Technology* 2020;15:561–9. <https://doi.org/10.1007/s42835-020-00346-4>.
- [134] Wang H, Zhang N, Du E, Yan J, Han S, Liu Y. A comprehensive review for wind, solar, and electrical load forecasting methods. *Global Energy Interconnection* 2022;5:9–30. <https://doi.org/10.1016/J.GLOEI.2022.04.002>.
- [135] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: A review. *Energy Convers Manag* 2018;156:459–97. <https://doi.org/10.1016/J.ENCONMAN.2017.11.019>.
- [136] Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews* 2020;124:109792. <https://doi.org/10.1016/J.RSER.2020.109792>.

- [137] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. *Renewable and Sustainable Energy Reviews* 2022;154:111768. <https://doi.org/10.1016/J.RSER.2021.111768>.
- [138] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renewable and Sustainable Energy Reviews* 2022;161:112348. <https://doi.org/10.1016/J.RSER.2022.112348>.
- [139] Jabir HJ, Teh J, Ishak D, Abunima H. Impacts of demand-side management on electrical power systems: A review. *Energies (Basel)* 2018;11. <https://doi.org/10.3390/en11051050>.
- [140] Dranka GG, Ferreira P. Review and assessment of the different categories of demand response potentials. *Energy* 2019;179:280–94. <https://doi.org/10.1016/J.ENERGY.2019.05.009>.
- [141] Iqbal S, Sarfraz M, Ayyub M, Tariq M, Chakraborty RK, Ryan MJ, et al. A Comprehensive Review on Residential Demand Side Management Strategies in Smart Grid Environment. *Sustainability* 2021, Vol 13, Page 7170 2021;13:7170. <https://doi.org/10.3390/SU13137170>.
- [142] Harper M. Review of Strategies and Technologies for Demand-Side Management on Isolated Mini-Grids 2013. <https://doi.org/10.2172/1171615>.
- [143] Kromer M, Zeifman M, Roth K, Yip T. The SunDial Framework: Enabling High Penetration Solar through the Integration of Energy Storage, Demand Management, and Forecasting. 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), IEEE; 2018.
- [144] Boait PJ, Snape JR. Demand management for isolated mini-grids supplied by renewable generation. *Proceedings of 2014 3rd International Conference on the Developments in Renewable Energy Technology, ICDRET 2014* 2014. <https://doi.org/10.1109/ICDRET.2014.6861688>.

- [145] Premkumar R, Shanmugasundaram R, Thilak KR, Balaji ANVSS. A Review on Demand Side Management: Definition, Scope, Challenges and Benefits. 2022 8th International Conference on Smart Structures and Systems (ICSSS) 2022:1–4. <https://doi.org/10.1109/ICSSS54381.2022.9782161>.
- [146] Mehra V, Amatya R, Ram RJ. Estimating the value of demand-side management in low-cost, solar micro-grids. *Energy* 2018;163:74–87. <https://doi.org/10.1016/J.ENERGY.2018.07.204>.
- [147] Shakeri M, Shayestegan M, Reza SMS, Yahya I, Bais B, Akhtaruzzaman M, et al. Implementation of a novel home energy management system (HEMS) architecture with solar photovoltaic system as supplementary source. *Renew Energy* 2018;125:108–20. <https://doi.org/10.1016/J.RENENE.2018.01.114>.
- [148] Parrish B, Heptonstall P, Gross R, Sovacool BK. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. *Energy Policy* 2020;138:111221. <https://doi.org/10.1016/J.ENPOL.2019.111221>.
- [149] Silva C, Faria P, Vale Z, Corchado JM. Demand response performance and uncertainty: A systematic literature review. *Energy Strategy Reviews* 2022;41:100857. <https://doi.org/10.1016/J.ESR.2022.100857>.
- [150] Li Z, Wang X, Lin D, Zheng R, Han B, Li G. Smart energy management model for households considering incentive-based demand response. *Proceedings - 2021 Power System and Green Energy Conference, PSGEC 2021* 2021:327–32. <https://doi.org/10.1109/PSGEC51302.2021.9541695>.
- [151] Hossain MM, Zafreen KR, Aziz T, Zamee MA. A price based demand side management strategy for residential sector in Bangladesh. *4th International Conference on Electrical Engineering and Information and Communication Technology, ICEEICT* 2018 2019:401–5. <https://doi.org/10.1109/CEEICT.2018.8628154>.
- [152] Alasseri R, Tripathi A, Joji Rao T, Sreekanth KJ. A review on implementation strategies for demand side management (DSM) in Kuwait

through incentive-based demand response programs. *Renewable and Sustainable Energy Reviews* 2017;77:617–35. <https://doi.org/10.1016/J.RSER.2017.04.023>.

[153] Park SC, Jin YG, Song HY, Yoon YT. Designing a critical peak pricing scheme for the profit maximization objective considering price responsiveness of customers. *Energy* 2015;83:521–31. <https://doi.org/10.1016/J.ENERGY.2015.02.057>.

[154] Zhang Y, Islam MM, Sun Z, Yang S, Dagli C, Xiong H. Optimal sizing and planning of onsite generation system for manufacturing in Critical Peaking Pricing demand response program. *Int J Prod Econ* 2018;206:261–7. <https://doi.org/10.1016/J.IJPE.2018.10.011>.

[155] Song H, Yoon YT, Kwon SC. Optimal scheduling of critical peak pricing considering photovoltaic generation and electric vehicle load. *ITEC Asia-Pacific 2019 - 2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific: New Paradigm Shift, Sustainable E-Mobility* 2019. <https://doi.org/10.1109/ITEC-AP.2019.8903840>.

[156] Ansarin M, Ghiassi-Farrokhfal Y, Ketter W, Collins J. A review of equity in electricity tariffs in the renewable energy era. *Renewable and Sustainable Energy Reviews* 2022;161:112333. <https://doi.org/10.1016/J.RSER.2022.112333>.

[157] Casini M. Smart buildings and smart cities. *Construction* 40 2022:607–60. <https://doi.org/10.1016/B978-0-12-821797-9.00012-X>.

[158] Fell MJ, Shipworth D, Huebner GM, Elwell CA. Public acceptability of domestic demand-side response in Great Britain: The role of automation and direct load control. *Energy Res Soc Sci* 2015;9:72–84. <https://doi.org/10.1016/J.ERSS.2015.08.023>.

[159] Wu YK, Tang KT. Frequency Support by Demand Response – Review and Analysis. *Energy Procedia* 2019;156:327–31. <https://doi.org/10.1016/J.EGYPRO.2018.11.150>.

- [160] Kalair AR, Abas N, Hasan QU, Seyedmahmoudian M, Khan N. Demand side management in hybrid rooftop photovoltaic integrated smart nano grid. *J Clean Prod* 2020;258:120747. <https://doi.org/10.1016/J.JCLEPRO.2020.120747>.
- [161] Paterakis NG, Erdinç O, Catalão JPS. An overview of Demand Response: Key-elements and international experience. *Renewable and Sustainable Energy Reviews* 2017;69:871–91. <https://doi.org/10.1016/J.RSER.2016.11.167>.
- [162] Alasseri R, Rao TJ, Sreekanth KJ. Institution of incentive-based demand response programs and prospective policy assessments for a subsidized electricity market. *Renewable and Sustainable Energy Reviews* 2020;117:109490. <https://doi.org/10.1016/J.RSER.2019.109490>.
- [163] Xiao D, do Prado JC, Qiao W. Optimal joint demand and virtual bidding for a strategic retailer in the short-term electricity market. *Electric Power Systems Research* 2021;190:106855. <https://doi.org/10.1016/J.EPSR.2020.106855>.
- [164] Utama C, Troitzsch S, Thakur J. Demand-side flexibility and demand-side bidding for flexible loads in air-conditioned buildings. *Appl Energy* 2021;285:116418. <https://doi.org/10.1016/J.APENERGY.2020.116418>.
- [165] Ribó-Pérez D, Carrión A, Rodríguez García J, Álvarez Bel C. Ex-post evaluation of Interruptible Load programs with a system optimisation perspective. *Appl Energy* 2021;303:117643. <https://doi.org/10.1016/J.APENERGY.2021.117643>.
- [166] Monyei CG, Adewumi AO. Integration of demand side and supply side energy management resources for optimal scheduling of demand response loads – South Africa in focus. *Electric Power Systems Research* 2018;158:92–104. <https://doi.org/10.1016/J.EPSR.2017.12.033>.
- [167] Li B, Liu Z, Wu Y, Wang P, Liu R, Zhang L. Review on photovoltaic with battery energy storage system for power supply to buildings: Challenges and opportunities. *J Energy Storage* 2023;61:106763. <https://doi.org/10.1016/J.EST.2023.106763>.

- [168] Trevizan RD, Headley AJ, Geer R, Atcitty S, Gyuk I. Integration of energy storage with diesel generation in remote communities. *MRS Energy & Sustainability* 2021 8:2 2021;8:57–74. <https://doi.org/10.1557/S43581-021-00013-9>.
- [169] Mohammad-Alikhani A, Mahmoudi A, Kahourzade S. Multi-objective optimization of system configuration and component capacity in AC mini-grid hybrid power system. 9th IEEE International Conference on Power Electronics, Drives and Energy Systems, PEDES 2020 2020. <https://doi.org/10.1109/PEDES49360.2020.9379526>.
- [170] Tsai CT, Beza TM, Molla EM, Kuo CC. Analysis and Sizing of Mini-Grid Hybrid Renewable Energy System for Islands. *IEEE Access* 2020;8:70013–29. <https://doi.org/10.1109/ACCESS.2020.2983172>.
- [171] Murugaperumal K, Ajay D Vimal Raj P. Feasibility design and techno-economic analysis of hybrid renewable energy system for rural electrification. *Solar Energy* 2019;188:1068–83. <https://doi.org/10.1016/J.SOLENER.2019.07.008>.
- [172] Rangel N, Li H, Aristidou P. An optimisation tool for minimising fuel consumption, costs and emissions from Diesel-PV-Battery hybrid microgrids. *Appl Energy* 2023;335:120748. <https://doi.org/10.1016/J.APENERGY.2023.120748>.
- [173] Zeynali S, Rostami N, Ahmadian A, Elkamel A. Two-stage stochastic home energy management strategy considering electric vehicle and battery energy storage system: An ANN-based scenario generation methodology. *Sustainable Energy Technologies and Assessments* 2020;39:100722. <https://doi.org/10.1016/J.SETA.2020.100722>.
- [174] Song Z, Guan X, Cheng M. Multi-objective optimization strategy for home energy management system including PV and battery energy storage. *Energy Reports* 2022;8:5396–411. <https://doi.org/10.1016/J.EGYR.2022.04.023>.

- [175] Kusakaka K, Phiri SF, Numbi BP. Optimal energy management of a hybrid diesel generator and battery supplying a RTG crane with energy recovery capability. *Energy Reports* 2021;7:4769–78. <https://doi.org/10.1016/J.EGYR.2021.07.074>.
- [176] Swetha MT, Nikhil S, Lakshmi PM, Mounika T, Girish Kumar G, Harsha Y, et al. Solar Based Battery Charging System Through IOT. *International Journal of Research in Engineering and Science (IJRES)* ISSN n.d.;10:1225–32.
- [177] Diouf B, Pode R. Potential of lithium-ion batteries in renewable energy. *Renew Energy* 2015;76:375–80. <https://doi.org/10.1016/J.RENENE.2014.11.058>.
- [178] Hariprakash B, Martha SK, Ambalavanan S, Gaffoor SA, Shukla AK. Comparative study of lead-acid batteries for photovoltaic stand-alone lighting systems. *J Appl Electrochem* 2008;38:77–82. <https://doi.org/10.1007/S10800-007-9403-4/FIGURES/5>.
- [179] Moita R, Monte D. The limits in the adoption of batteries. *Energy Econ* 2022;107:105776. <https://doi.org/10.1016/J.ENERCO.2021.105776>.
- [180] Pearre N, Swan L. Reimagining renewable electricity grid management with dispatchable generation to stabilize energy storage. *Energy* 2020;203:117917. <https://doi.org/10.1016/J.ENERGY.2020.117917>.
- [181] Canizes B, Silveira V, Vale Z. Demand response and dispatchable generation as ancillary services to support the low voltage distribution network operation. *Energy Reports* 2022;8:7–15. <https://doi.org/10.1016/J.EGYR.2022.01.040>.
- [182] Bird L, Lew D, Milligan M, Carlini EM, Estanqueiro A, Flynn D, et al. Wind and solar energy curtailment: A review of international experience. *Renewable and Sustainable Energy Reviews* 2016;65:577–86. <https://doi.org/10.1016/J.RSER.2016.06.082>.

- [183] López Prol J, Zilberman D. No alarms and no surprises: Dynamics of renewable energy curtailment in California. *Energy Econ* 2023;126:106974. <https://doi.org/10.1016/J.ENERCO.2023.106974>.
- [184] Rezaeimozafer M, Monaghan RFD, Barrett E, Duffy M. A review of behind-the-meter energy storage systems in smart grids. *Renewable and Sustainable Energy Reviews* 2022;164:112573. <https://doi.org/10.1016/J.RSER.2022.112573>.
- [185] Ahammed MT, Khan I. Ensuring power quality and demand-side management through IoT-based smart meters in a developing country. *Energy* 2022;250:123747. <https://doi.org/10.1016/J.ENERGY.2022.123747>.
- [186] Li W, Logenthiran T, Phan VT, Woo WL. Intelligent housing development building management system (HDBMS) for optimized electricity bills. *Conference Proceedings - 2017 17th IEEE International Conference on Environment and Electrical Engineering and 2017 1st IEEE Industrial and Commercial Power Systems Europe, IEEEIC / I and CPS Europe 2017* 2017. <https://doi.org/10.1109/IEEEIC.2017.7977410>.
- [187] Gunasekara R, Meegahapola L, Jalili M. Demand Response Scheme for Distribution Transformer Load Relieving using Smart-Meter Data. *Proceedings of 2021 31st Australasian Universities Power Engineering Conference, AUPEC 2021* 2021. <https://doi.org/10.1109/AUPEC52110.2021.9597818>.
- [188] Mlynek P, Uher V, Misurec J. Forecasting of Smart Meters Energy Consumption for Data Analytics and Grid Monitoring. *2022 22nd International Scientific Conference on Electric Power Engineering (EPE)* 2022:1–5. <https://doi.org/10.1109/EPE54603.2022.9814101>.
- [189] Chakravarthi PK, Yuvaraj D, Venkataramanan V. IoT-based smart energy meter for smart grids. *ICDCS 2022 - 2022 6th International Conference on Devices, Circuits and Systems* 2022:360–3. <https://doi.org/10.1109/ICDCS54290.2022.9780714>.

- [190] Massaferro P, Martino JM di, Fernandez A. Fraud Detection on Power Grids While Transitioning to Smart Meters by Leveraging Multi-Resolution Consumption Data. *IEEE Trans Smart Grid* 2022;13:2381–9. <https://doi.org/10.1109/TSG.2022.3148817>.
- [191] Avancini DB, Rodrigues JJPC, Martins SGB, Rabêlo RAL, Al-Muhtadi J, Solic P. Energy meters evolution in smart grids: A review. *J Clean Prod* 2019;217:702–15. <https://doi.org/10.1016/J.JCLEPRO.2019.01.229>.
- [192] Farinis G, Kanellos FD. Integrated energy management system for Microgrids of building prosumers. *Electric Power Systems Research* 2021;198:107357. <https://doi.org/10.1016/J.EPSR.2021.107357>.
- [193] Murugaperumal K, Raj PADV. Integrated energy management system employing pre-emptive priority based load scheduling (PEPLS) approach at residential premises. *Energy* 2019;186:115815. <https://doi.org/10.1016/J.ENERGY.2019.07.145>.
- [194] Jabbour N, Tsioumas E, Papagiannis D, Koseoglou M, Mademlis C. An Integrated Energy Management System for Nearly Zero Energy Buildings. *ICCEP 2019 - 7th International Conference on Clean Electrical Power: Renewable Energy Resources Impact* 2019:259–65. <https://doi.org/10.1109/ICCEP.2019.8890174>.
- [195] Wang X, Li Y, Wang C, Lin H, Li M. Research on Integrated Energy Management System Based on Microservice Architecture. *Proceedings - 2021 International Conference on Power System Technology: Carbon Neutrality and New Type of Power System, POWERCON 2021* 2021:401–6. <https://doi.org/10.1109/POWERCON53785.2021.9697774>.
- [196] Zhang Y, Vand B, Baldi S. A Review of Mathematical Models of Building Physics and Energy Technologies for Environmentally Friendly Integrated Energy Management Systems. *Buildings* 2022;12. <https://doi.org/10.3390/BUILDINGS12020238>.

- [197] Yi L, Langlotz P, Hussong M, Glatt M, Sousa FJP, Aurich JC. An integrated energy management system using double deep Q-learning and energy storage equipment to reduce energy cost in manufacturing under real-time pricing condition: A case study of scale-model factory. *CIRP J Manuf Sci Technol* 2022;38:844–60. <https://doi.org/10.1016/J.CIRPJ.2022.07.009>.
- [198] Wang Y, Qin Y, Ma Z, Wang Y, Li Y. Operation optimisation of integrated energy systems based on cooperative game with hydrogen energy storage systems. *Int J Hydrogen Energy* 2023;48:37335–54. <https://doi.org/10.1016/J.IJHYDENE.2023.06.170>.
- [199] Hengrui M, Wang B, Wang H, Wang S, Fuqi M, Wang X, et al. An effective planning approach for integrated energy systems considering equipment operating characteristics. *Heliyon* 2023;9:e21409. <https://doi.org/10.1016/J.HELİYON.2023.E21409>.
- [200] Guo Y, Xiang Y. IESP: An integrated energy system planning platform. *Energy Reports* 2023;9:242–8. <https://doi.org/10.1016/J.EGYR.2023.04.280>.
- [201] Huang S, Lu H, Chen M, Zhao W. Integrated energy system scheduling considering the correlation of uncertainties. *Energy* 2023;283:129011. <https://doi.org/10.1016/J.ENERGY.2023.129011>.
- [202] Feng B, Fu Y, Huang Q, Ma C, Sun Q, Wennersten R. Multi-objective optimization of an integrated energy system with high proportion of renewable energy under multiple uncertainties. *Energy Reports* 2023;9:695–701. <https://doi.org/10.1016/J.EGYR.2023.05.194>.
- [203] He C, Zhu J, Cheung K, Borghetti A, Zhang D, Zhu H. Optimal planning of integrated energy system considering swapping station and carbon capture power system. *Energy Reports* 2023;9:165–70. <https://doi.org/10.1016/J.EGYR.2023.09.120>.
- [204] Papadopoulos CEm. Multi-Objective Decision Optimization in PLEXOS® Integrated Energy Models. *Energy Exemplar* 2023.

<https://www.energyexemplar.com/multi-objective-decision-optimization>
(accessed November 28, 2023).

[205] Ringkjøb HK, Haugan PM, Solbrekke IM. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews* 2018;96:440–59. <https://doi.org/10.1016/J.RSER.2018.08.002>.

[206] Khalili S, Breyer C. Review on 100% Renewable Energy System Analyses - A Bibliometric Perspective. *IEEE Access* 2022;10:125792–834. <https://doi.org/10.1109/ACCESS.2022.3221155>.

[207] Lopez G, Aghahosseini A, Child M, Khalili S, Fasihi M, Bogdanov D, et al. Impacts of model structure, framework, and flexibility on perspectives of 100% renewable energy transition decision-making. *Renewable and Sustainable Energy Reviews* 2022;164:112452. <https://doi.org/10.1016/J.RSER.2022.112452>.

[208] Groppi D, Feijoo F, Pfeifer A, Garcia DA, Duic N. Analyzing the impact of demand response and reserves in islands energy planning. *Energy* 2023;278:127716. <https://doi.org/10.1016/J.ENERGY.2023.127716>.

[209] Ren GW, Yu M, Yin DS, Huang SY, Xu HC, Yuan M. Design and optimization of integrated energy management network system based on internet of things technology. *Sustainable Computing: Informatics and Systems* 2021;30:100502. <https://doi.org/10.1016/J.SUSCOM.2020.100502>.

[210] Huang W, Zhang N, Kang C, Li M, Huo M. From demand response to integrated demand response: review and prospect of research and application. *Protection and Control of Modern Power Systems* 2019;4:1–13. <https://doi.org/10.1186/S41601-019-0126-4/TABLES/1>.

[211] Deka K, Mahapatra S, Bordoloi PK, Medhi B. Design and testing of a control system for supply side electrical load management. *Proceedings of 2014 1st International Conference on Non-conventional Energy: Search for Clean and*

Safe Energy, ICONCE 2014 2014:277–81.
<https://doi.org/10.1109/ICONCE.2014.6808726>.

[212] Luo XJ, Fong KF. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. *Appl Energy* 2019;242:570–87.
<https://doi.org/10.1016/J.APENERGY.2019.03.149>.

[213] Ghiasi MI, Hajizadeh A, Golkar MA, Marefati M. Demand and supply side management strategies for zero energy buildings. 2017 IEEE 17th International Conference on Ubiquitous Wireless Broadband, ICUWB 2017 - Proceedings 2018;2018-January:1–5. <https://doi.org/10.1109/ICUWB.2017.8251007>.

[214] Ortiz JM, Kracht W, Pamparana G, Haas J. Optimization of a SAG Mill Energy System: Integrating Rock Hardness, Solar Irradiation, Climate Change, and Demand-Side Management. *Mathematical Geosciences* 2019 52:3 2019;52:355–79. <https://doi.org/10.1007/S11004-019-09816-6>.

[215] Fujimoto Y, Kikusato H, Yoshizawa S, Kawano S, Yoshida A, Wakao S, et al. Distributed energy management for comprehensive utilization of residential photovoltaic outputs. *IEEE Trans Smart Grid* 2018;9:1216–27. <https://doi.org/10.1109/TSG.2016.2581882>.

[216] Aslam S, Khalid A, Javaid N. Towards efficient energy management in smart grids considering microgrids with day-ahead energy forecasting. *Electric Power Systems Research* 2020;182:106232. <https://doi.org/10.1016/J.EPSR.2020.106232>.

[217] Sharda S, Sharma K, Singh M. A real-time automated scheduling algorithm with PV integration for smart home prosumers. *Journal of Building Engineering* 2021;44:102828. <https://doi.org/10.1016/J.JOBE.2021.102828>.

[218] Manur A, Marathe M, Manur A, Ramachandra A, Subbarao S, Venkataramanan G. Smart Solar Home System with Solar Forecasting. 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable

Energy, PESGRE 2020 2020.
<https://doi.org/10.1109/PESGRE45664.2020.9070340>.

[219] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. *Solar Energy* 2016;136:78–111. <https://doi.org/10.1016/J.SOLENER.2016.06.069>.

[220] Papadimitriou CN, Anastasiadis A, Psomopoulos CS, Vokas G. Demand Response schemes in Energy Hubs: A comparison study. *Energy Procedia* 2019;157:939–44. <https://doi.org/10.1016/J.EGYPRO.2018.11.260>.

[221] Nicolson ML, Fell MJ, Huebner GM. Consumer demand for time of use electricity tariffs: A systematized review of the empirical evidence. *Renewable and Sustainable Energy Reviews* 2018;97:276–89. <https://doi.org/10.1016/J.RSER.2018.08.040>.

[222] Tabares-Velasco PC, Speake A, Harris M, Newman A, Vincent T, Lanahan M. A modeling framework for optimization-based control of a residential building thermostat for time-of-use pricing. *Appl Energy* 2019;242:1346–57. <https://doi.org/10.1016/J.APENERGY.2019.01.241>.

[223] Carmichael R, Gross R, Hanna R, Rhodes A, Green T. The Demand Response Technology Cluster: Accelerating UK residential consumer engagement with time-of-use tariffs, electric vehicles and smart meters via digital comparison tools. *Renewable and Sustainable Energy Reviews* 2021;139:110701. <https://doi.org/10.1016/J.RSER.2020.110701>.

[224] Qiu Y, Kirkeide L, Wang YD. Effects of Voluntary Time-of-Use Pricing on Summer Electricity Usage of Business Customers. *Environ Resour Econ (Dordr)* 2018;69:417–40. <https://doi.org/10.1007/S10640-016-0084-5/TABLES/10>.

[225] Nourollahi R, Zare K, Nojavan S. Energy management of hybrid AC-DC microgrid under demand response programs: Real-time pricing versus timeof-use pricing. *Demand Response Application in Smart Grids: Operation Issues - Volume 2* 2019:75–93. https://doi.org/10.1007/978-3-030-32104-8_4/COVER.

- [226] Amin A, Tareen WUK, Usman M, Ali H, Bari I, Horan B, et al. A Review of Optimal Charging Strategy for Electric Vehicles under Dynamic Pricing Schemes in the Distribution Charging Network. *Sustainability* 2020, Vol 12, Page 10160 2020;12:10160. <https://doi.org/10.3390/SU122310160>.
- [227] Cui Q, Wang X, Wang X, Zhang Y. Residential Appliances Direct Load Control in Real-Time Using Cooperative Game. *IEEE Transactions on Power Systems* 2016;31. <https://doi.org/10.1109/TPWRS.2015.2391774>.
- [228] Rastegar M, Fotuhi-Firuzabad M, Lehtonen M. Home load management in a residential energy hub. *Electric Power Systems Research* 2015;119:322–8. <https://doi.org/10.1016/J.EPSR.2014.10.011>.
- [229] IRENA. Innovation landscape brief: Time-of-use tariffs. International Renewable Energy Agency 2019:1–184. www.irena.org (accessed October 19, 2022).
- [230] Kiguchi Y, Weeks M, Arakawa R. Predicting winners and losers under time-of-use tariffs using smart meter data. *Energy* 2021;236:121438. <https://doi.org/10.1016/J.ENERGY.2021.121438>.
- [231] Hafiz F, Awal MA, De Queiroz AR, Husain I. Real-Time Stochastic Optimization of Energy Storage Management Using Deep Learning-Based Forecasts for Residential PV Applications. *IEEE Trans Ind Appl* 2020;56:2216–26. <https://doi.org/10.1109/TIA.2020.2968534>.
- [232] Goodarzi S, Perera HN, Bunn D. The impact of renewable energy forecast errors on imbalance volumes and electricity spot prices. *Energy Policy* 2019;134:110827. <https://doi.org/10.1016/J.ENPOL.2019.06.035>.
- [233] Jiang Q, Xue M, Geng G. Energy management of microgrid in grid-connected and stand-alone modes. *IEEE Transactions on Power Systems* 2013;28:3380–9. <https://doi.org/10.1109/TPWRS.2013.2244104>.
- [234] Nge CL, Ranaweera IU, Midtgård OM, Norum L. A real-time energy management system for smart grid integrated photovoltaic generation with

battery storage. *Renew Energy* 2019;130:774–85.
<https://doi.org/10.1016/J.RENENE.2018.06.073>.

[235] Elkazaz M, Sumner M, Naghiyev E, Pholboon S, Davies R, Thomas D. A hierarchical two-stage energy management for a home microgrid using model predictive and real-time controllers. *Appl Energy* 2020;269:115118.
<https://doi.org/10.1016/J.APENERGY.2020.115118>.

[236] Fang Y, Zhao S, Chen Z. Multi-objective unit commitment of jointly concentrating solar power plant and wind farm for providing peak-shaving considering operational risk. *International Journal of Electrical Power & Energy Systems* 2022;137:107754. <https://doi.org/10.1016/J.IJEPES.2021.107754>.

[237] Hytowitz RB, Hedman KW. Managing solar uncertainty in microgrid systems with stochastic unit commitment. *Electric Power Systems Research* 2015;119:111–8. <https://doi.org/10.1016/J.EPSR.2014.08.020>.

[238] Wang K, Xie Y, Zhang W, Cai H, Liang F, Li Y. Research on optimal dispatch of distributed energy considering new energy consumption. *Energy Reports* 2023;10:1888–98. <https://doi.org/10.1016/J.EGYR.2023.08.040>.

[239] Aktar AK, Taşcıkaraoğlu A, Gürleyük SS, Catalão JPS. A framework for dispatching of an electric vehicle fleet using vehicle-to-grid technology. *Sustainable Energy, Grids and Networks* 2023;33:100991.
<https://doi.org/10.1016/J.SEGAN.2022.100991>.

3 A Three-step weather data approach in solar energy prediction using machine learning²

Abstract

Solar energy plays a critical part in lowering CO₂ emissions and other greenhouse gases when integrated into the grid. Higher solar energy penetration is hindered by its intermittency leading to reliability issues. To forecast solar energy production, this study proposes a three-step forecasting method that selects weather variables with a moderate to strong positive correlation to solar radiation. Using low-level data fusion to combine weather inputs from a reliable local weather station and an on-site weather station, significantly improves the forecasting model's accuracy regardless of the machine learning method used. Weather data was obtained from the Kisanhub Weather Station located in Cranfield University, UK and the meteorological station in Bedford, UK. In addition, PV power supply data was obtained from four solar plants. Using the Regression Learner app in MATLAB, the proposed architecture is tested on a utility scale solar plant (1 MW), showing a 6% and 13% prediction accuracy improvement when compared to solely using data from the on-site and local weather station respectively. It is further validated using data from three residential rooftop solar systems (8 kW, 10.5 kW and 15 kW), achieving root-mean square values of 0.0859, 0.0984, 0.1425 and 0.0885 respectively.

Keywords: variable renewable energy, solar energy forecasting, machine learning, solar radiation, low-level data fusion

² This chapter is a paper authored by Tolulope Falope, Liyun Lao, & Dawid Hanak, submitted to the Renewable Energy Focus journal.

3.1 Introduction

Recent trends show an uptake in the adoption of renewable energy (RE) technologies resulting in the greening of the grid. However, the intermittency of some renewable energy sources (RES) presents a challenge to grid managers because occasionally energy supply does not match up with demand [1,2]. Increasing RE penetration levels mean system operators and prosumers must strike a balance between energy generation and consumption to preserve grid stability. Grid integration involves increasing the share of variable renewable energy (VRE) in the grid using effective and cost-effective strategies, while preserving or enhancing system stability and reliability [3].

RES like solar and wind are abundant in nature and are already widely to meet global energy demands. Integrating solar generation into the grid depends on its accurate prediction [4]. Proper system planning and grid integration is impossible without improving forecasting [5].

Solar forecasting involves gathering and analyzing data to predict future solar irradiance or solar power [6]. Data can be historical or in real-time, and can be provided by weather stations, remote sensing (satellite imaging) and so on.

While choosing a solar forecasting strategy, there are five main factors to consider. These include the input or raw data's source and format, data mapping or the forecasting architecture, the forecasting methodology, the desired output or targeted outcome, and the forecasting time frame.

The first factor deals with the type of weather data, categorized into historical, real-time, and forecast data. In addition, the data source is also important. Example of data sources include on-site and local weather outposts, or even software generated or simulated data from sources like MATLAB.

The second factor is to create a model that describes the relationship between input variables and output variables. [6]. The strength of the prediction model is reliant on the correlation of input variable(s) to the output [7].

The third factor is the forecasting method. Solar forecasting methods are classified into physical and statistical (or non-physical) methods. Physical methods predict outputs like solar irradiation by using numerical weather predictions (NWP) models with weather variable inputs like temperature, relative humidity, cloud cover and so on [8,9]. Statistical methods depend purely on historical and/or real-time data to arrive at predicted variable(s) that have a linear correlation with the data over the same time frame. Hybrid models, which offer a higher predicting accuracy, can be created by combining several methods or forecasting algorithms.

The fourth factor is the predicted or output variable, which categorizes forecasting into direct (solar power generated) or indirect (solar irradiance) [7]. Direct forecasting involves forecasting the PV power output directly [10]. Indirect forecasting predicts the solar irradiation first before estimating the PV output using mathematical formulas that describes the specific system under consideration. The solar power and the PV generated are one and the same thing.

The last factor is the forecasting time horizon or timescale. This is the time frame from the present into the future for which an outcome is predicted. Forecasting time horizons are classified into intra-hour or very short-term (5 mins to 60 mins), short-term (1 hour to 6 hours), medium-term (1 day to several days), long-term (week(s) to a year or more). Each of the time horizons is suitable for specific applications. For example, very short-term forecasting is used to achieve responsive control for RE power plants, market clearing, real-time dispatch, and load tracking. Short-term forecasting is used to schedule energy flow from power sources, load-following, and initiating back-up storage. Medium-term forecasting is used for congestion management, market trading and price settlement, predicting reserve requirements and load dispatch. Long-term forecasting is used for contingency analysis, operation management, and resource and maintenance planning [8,11].

Based on the above considerations, several studies have developed varying approaches to forecast solar energy supply with an emphasis on a next

day or 24-hour ahead (day-ahead) timescale. In Carrera et al. [12], a combination of forecast and historical weather data from an on-site and a local weather station was used, respectively, as the data source in addition to historical PV generation and sun elevation. Minhas et al. [13] uses historical and forecast data from a local weather station and the historical PV data from the solar plant. Table 3-1 gives a summary of forecasting architecture for day-ahead PV generated predictions.

Table 3-1 Summary of day-ahead solar energy forecasting studies

Research study	Historical observed weather data from weather station	Historical forecast weather data	Historical observed weather data on-site	Historical observed PV data
This study	✓		✓	✓
Carrera et al. [12]	✓	✓		✓
Minhas et al. [13]			✓	✓
Aghajani et al. [14]		✓		✓
Manjarres et al. [15]	✓	✓		✓
Pawar et al. [16]	✓			✓
Sangrody et al. [17]	✓			✓
Pawar et al. [18]	✓			✓
Feng and Yu [19]				✓

As can be seen by the summary given in Table 1, this study proposes a unique combination of historical photovoltaic (PV) data and historical weather data from both an on-site and local weather station to forecast the next-day PV supply. None of the other studies has this data combination. This is significant for a few reasons. The first reason is that the combined sum of data from multiple relevant data sources is better and richer than information from any one individual source; obtaining more effective information and optimizing data combination [20]. The two key words are multiple and relevant. The better the data quality, the higher the accuracy. This study combines data coming from two relevant independent sources (local and an on-site weather station) using data fusion. The aim of data fusion is to improve the overall accuracy of a system or model by integrating data from multiple sources. Data fusion can be categorized into low-level, mid-level or intermediate, and high-level fusion. Low-level data fusion (LLDF) or

measurement level fusion [21] combines more than one source of raw data to produce new raw data [22].

The proposed forecasting architecture is a three-step process that predicts PV generation by selecting weather variables with moderate to strong positive correlations to solar radiation (SR) from the two weather stations mentioned above, combined with historical PV generation data obtained from a PV power plant, into machine learning algorithms from MATLAB's regression learner app (RLA). SR is the most important predictor to solar power, or PV generated [12,23,24] hence its selection as the reference parameter. Additionally, the correlation between the input and output model variables has a significant impact on how well a forecasting model performs [7]. With the output variable being the PV generated, it is logical to choose inputs that have a strong correlation to PV generation. To investigate the accuracy of the proposed architecture, it is applied to both commercial/utility and residential solar plants.

The MATLAB RLA is a machine learning toolbox app used to train forecasting algorithms based on known output and input parameters. It selects the best performing out of more than twenty algorithms based on the lowest root mean square error (RMSE).

3.2 Data source and type: A Literature review

The input data can be made up of either historical weather data (observed and/or forecast), real-time data, historical solar energy generation data, or time of day data. While a few studies employ a univariate approach where only the historical PV generation is used [25–28], this research finds that most of the previous studies rely on two sources of data: the historical PV generation and some form of weather data [12–15,17,18,29–54]. This may be because multivariate approaches perform better than univariate [55]. Fewer studies use more than two data sources. This is important because data source is critical to the model performance [50]. In addition, not all site locations have access to both sets of data. Some sites only receive forecast data while others only have observed data.

A further dimension in evaluating the input weather data used is the weather data proximity to the generation plant. The solar energy production of a solar distributed system, solar farm or mini grid depends on the local weather conditions, particularly the SR of that specific location. In fact, no two PV systems show the same effect from meteorological parameters due to their unique combination, tied to differing geographical locations [7]. In other words, no two locations will have the same correlation of meteorological parameters and PV power generated. In other words, there won't be a consistent correlation between meteorological factors and PV power generated at any two locations. Despite their interdependence, local, regional, and global meteorological data offer varying degrees of accuracy. Most research studies use data from local or regional weather stations for their weather analyses. Although these data sources can be very accurate, it is not uncommon that a mismatch exists between them and exact data from the site. Access to an on-site weather station improves the quality of weather data [56]. Site-specific conditions that affect the weather but are not represented in local or regional weather data can be recorded by an on-site weather station. To get the most accurate result, it is best to pair data from a local weather station with data from an on-site weather station where the forecasting system is being deployed [56]. Other options are to exclusively use the data from the on-site weather station. For instance, [29] uses a rooftop PV station located at the solar array site to collect all the data used by the study. [30–33,57] are studies that also use this approach.

The use of observed weather data from a local or regional weather station tends to be more accurate than forecast data because they include more variables which may have a direct correlation to solar generation [50]. For example, Ngoc-Lan Huynh et al [34] use solar radiation and Pedregal et al. [35] use global horizontal irradiation as the only weather parameters inputs into their prediction models. Other studies take a multivariate weather approach. Rodríguez et al. [39] use pressure, humidity, and temperature in their forecast model. Thukral et al. [40] (air temperature, relative humidity, wind, solar radiation), Chen et al. [44] (relative humidity, temperature, global horizontal radiation), Korkmaz [37] (solar radiation, temperature, humidity) and several

other studies [36,42,43,58–60] use this multivariate weather approach. Whether using one weather parameter or using several, observed data tends to be more accurate than forecast data. While this is good, using weather data from only one source can be problematic if the source is somehow compromised.

Iyengar et al. [45] uses historical forecast data as a primary predictor and is another approach used by other studies [46–48]. This data approach has an advantage that is also its Achilles heel. Forecast data needs to be accurate to inspire confidence in the data by the general public, who are the core users. The drawback is that the data then tends to be more conservative due to concerns of inaccuracy and can be a bit vague by giving categorical variables instead of numerical variables. For example, the United Kingdom's national Meteorological Office (the Met Office), gives a forecast in Bedford UK of <5% as the chance of precipitation instead of the actual percentage figure. Another drawback as mentioned earlier is forecast data tend to have fewer weather parameters than observed data. In the same Bedford weather station, the Met office has seven weather parameters it forecasts as opposed to twelve for observed data.

Carrera et al. [12], Kim et al. [50] and Abedinia et al. [49] use a hybrid approach, incorporating forecast data with observed data. This method is more accurate [51] because it incorporates the advantages of both sources.

In a two-stage methodology, Kim et al. [50] uses an auxiliary model in the first stage to identify the relationship between forecast and auxiliary variables in the weather data. In the second stage, the model receives the output from the first to forecast the PV energy. This results in an improvement in the forecasting model accuracy. Kyliashkina et al. [51] suggest an alternative strategy in which forecast data is adjusted using the observed data rather than the auxiliary variables. Abedinia et al. [49] selects both historical and forecast data for solar irradiation, weather temperature and cell temperature; three weather parameters with strong correlations to PV generation. However, local conditions specific to the site are not reflected in the forecast.

Given the analysis of the above data sources, a three-data source approach is proposed. This study will use historical observed data from both an

on-site and local weather station, and the historical PV data. It is intuitive that the level of accuracy of the weather data in this approach increases as it gets closer to the power generating site. In other words, the proximity of the weather station to the PV site is directly proportional to the accuracy of the weather data and must be considered when using data [39]. The reason for using the most accurate weather data is that the forecasting model's performance depends on the accuracy of the database [39]. In addition, the choice of the local weather station was made based on proximity to PV plant to minimize spatial variability, the granularity and completeness of the data, and the inclusion of key weather parameters in the database.

Of significance, is the novelty of this approach as none of the studies use LLDF with data from an on-site weather station and a trustworthy local weather station. Site-adaptation techniques have been developed to improve satellite-derived weather data by using in-situ data from meteorological stations [24]. But these methods rely on complex machine and deep learning algorithms to transform satellite images into solar irradiation values as an example. Jesus et al. [61] uses data fusion based on a hybrid of two deep neural networks namely Fully Convolutional Network (FCN) and Long Short-Term Memory (LSTM) network, to fuse the output of two individual forecast models known as Autoregressive Moving Average with Exogenous Inputs (ARMAX) and Adaptive Neuro Fuzzy Inference System (ANFIS) to predict short-term PV output. In another short-term prediction study, Wang et al. [20] uses multi-source data fusion to fuse NWP, PV output data and remote satellite sensing data with a Support Vector Machine (SVM) algorithm. Results showed that the data fused SVM model performed better than the regular SVM model more than 95% of the time.

3.3 Methods

The data and procedures used to develop the forecasting model are described in this section. The source and data type are described in Section 3.3.1. The correlation coefficient and matrix used to identify the most important

meteorological parameters are briefly explained in Section 3.3.2. Section 3.3.3 briefly describes the data pre-processing methods used to clean the data. The error equations used to assess the model performance are described in Section 3.3.4. The methodology and flow chart of the RLA in MATLAB are highlighted in Section 3.3.5.

3.3.1 Data description

In this case study, the Kisanhub weather station (KH075) [62] at Cranfield University, recorded data every fifteen minutes on nine weather parameters, including solar irradiation. Historical PV data containing half-hourly power generation from four PV plants located in Cranfield University representing the spectrum of decentralized distributed generation (DDG) customers is obtained from Cranfield's online energy manager platform [63]. The aim is to investigate the proposed architecture's performance in both commercial and residential applications. The specification and data profiles of the locations are listed in Table 3-2 and Table 3-3. 75% of the data from the 1 MW solar farm was used for training the model while the remaining 25% was used for testing. 100% of the data from the three residential solar arrays was used to validate the model. Upon request, The Metrological Office for Bedford, U.K., the station closest to Cranfield, provided hourly data for twelve weather parameters. Missing data points were simply deleted from the data sample.

Table 3-2 PV system specifications of 1 MW solar farm in Cranfield University

Location	1 MW solar farm, Cranfield University
PV Generator Output	999.78 kWp
PV Generator Surface	5,858.4 m ²
Number of PV Modules	3508
Type of PV Modules	Hanwha Q cells: 285 W
Number of Inverters	35
Type of Inverters	Fronius Symo: 10 kW to 17.5 kW
Data Sample Period	January 2019 – December 2020
Training (% of data sample)	75%
Testing (% of data sample)	25%

Table 3-3 PV system specification of three roof-top solar arrays in Cranfield University

Building no.	B052A	B146 FAAM	B252
PV Generator Output	10.45 kWp	15 kWp	8.26 kWp
Number of PV Modules	35	60	28
Type of PV Modules	BYD: 275 W	Vikram: 250 W	Q.antum: 295 W
Number of Inverters	1	1	1
Type of Inverter	Solis: 12 kW	PVI: 12.5 kW	Solar Edge: 10 kW
Data Sample Period	Jan. 2020 – Dec. 2020	July 2018 – Sept. 2021	Dec. 2019 – Oct. 2021

The varied data sample periods of each PV plant in **Error! Reference source not found.** and Table 3-3 were due to data availability on the energy manager portal.

3.3.2 Data comparison

The Pearson Correlation Coefficient (PCC) function [64] is given in Equation (3-1) and is used to determine which weather variable has a moderate or strong positive correlation to SR. Results are displayed in a correlation matrix [65].

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (3-1)$$

Where r is the correlation coefficient, x and y represent the values of the x-variable and y-variable in the sample respectively, $\bar{x}$ and $\bar{y}$ are the mean of the x-variable values and the y-variable values respectively.

The correlation coefficients, with a range of values between -1 to +1, indicates how strong the linear relationship between two variables is. A perfect negative correlation is equivalent to -1 on the scale while a perfect positive is equivalent to +1. This study is interested in weather parameters with a moderate to strong positive correlation to SR, which is equivalent to having a PCC greater than 0.4 [66]. The implication is that having a strong positive linear relationship to the most significant predictor of PV generation (SR) suggests the weather parameter is also a good predictor.

In addition, PCC helps to ensure data accuracy and reliability from both weather stations by cross validating the correlation coefficients of identical weather variables common to both weather stations.

3.3.3 Data pre-processing

Different weather parameters have varying units, scales and dimensions that can form a bias in determining the forecasting model. For this reason, it is imperative to normalize all variables to the same order of magnitude [67]. Re-scaling or the min-max nominalization method (RM) is a popular and widely used method for normalizing data [68]. In this study, the intervals [a b] for rescaling were selected as [0 1]. The RM equation [69] is given in Equation (3-2).

$$X_{rescaled} = a + \left[\frac{X - X_{min}}{X_{max} - X_{min}} \right] (b - a) \quad (3-2)$$

Where $X_{rescaled}$ is the normalized data, X represents the values of the x-variable in the sample, X_{min} and X_{max} are the minimum and maximum values of the x-variable in the sample respectively. To harmonize the timescales, the quarterly and half-hourly weather and PV data from Cranfield that do not match the hourly weather data from Bedford are ignored. In other words, only corresponding hourly data that is present across all the data sources are considered for nominalization.

3.3.4 Model performance evaluation

To evaluate the performance of the model under considered scenarios, four performance statistical error functions [70] and the coefficient of determination, R^2 [71], were used. The mean squared error (MSE) shown in Equation (3-3) measures the average of the square of the errors. The mean absolute error (MAE) described in Equation

(3-4) measures the average magnitude of the errors. Equation (3-5), which is the RMSE, is simply the square root of the MSE. R^2 in Equation (3-6) is the proportion of the change in the dependent variable that can be predicted due to the corresponding change in

independent variable(s). When evaluating how well the underlying models fit the data, this coefficient of determination is typically used because it is well-defined for linear regression models.

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2 = RMSE^2 \quad (3-3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |(\hat{x}_i - x_i)| \quad (3-4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (3-5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{x}_i - x_i)^2}{\sum_{i=1}^N (\hat{x}_i - \hat{x}_{average})^2} \quad (3-6)$$

Where N is the number of data points, $\hat{x}_i$ are the predicted values and x_i are the observed values.

3.3.5 Regression learner app methodology and flow chart

After pre-processing the data, it is then trained using the RLA; an application in the ML Toolbox in MATLAB. Regression analysis is used to determine the relationship and future predictions of a dependent variable (the PV generated) from a number of independent variables (weather parameters). The RLA is a useful tool because it allows the user to train and validate a model using more than twenty different forecasting algorithms. It has also been used successfully in numerous machine learning applications that require forecasting [72–77]. After training, the user can compare the algorithms side by side based on their errors and choose the best model. The app identifies the best forecasting algorithm

based on the lowest RMSE and highlights the model in the results display. Another advantage of the RLA is that multiple algorithms can be run simultaneously through a “parallel” function which allows the user to view the results in a shorter time frame than running each algorithm one after the other. To initiate the model training process via the app, the response (dependent variable) and predictors (independent variables) with the same corresponding time stamps must be selected. For this case study, the correlated weather data are the predictors while the response or output is the PV generation. **Error! Reference source not found.** highlights the flow chart of the RLA.

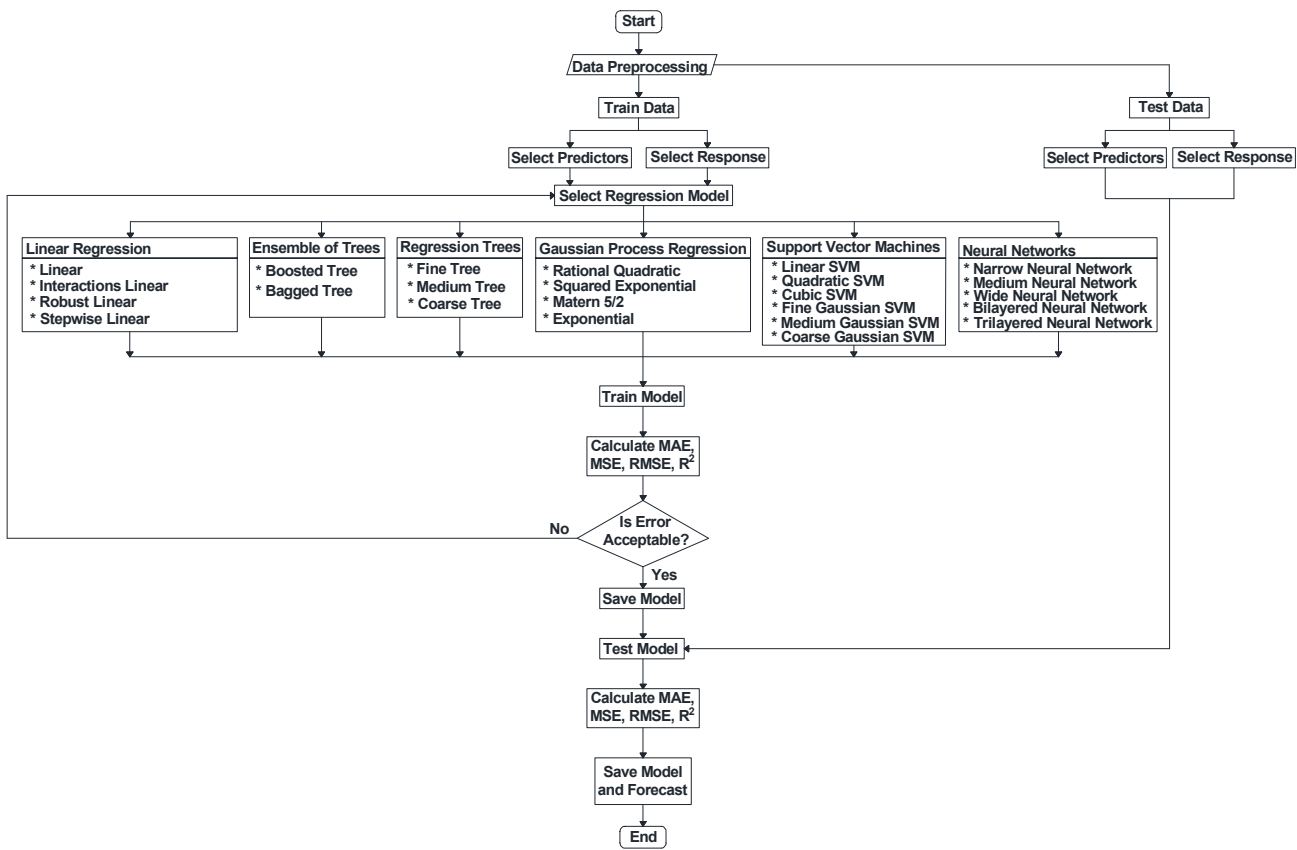


Figure 3-1 Flow chart of the regression learner app methodology

3.4 Methodology

The justification for employing the combined data from both weather stations is that weather stations capture weather parameters differently (different

methodologies) and often don't capture the same type or number of parameters. The premise in this section supports the assertion made above. The three-step forecasting architecture used to correlate input data with energy supply is described in Section 3.4.1. To evaluate the efficiency of the forecasting architecture, Section 3.4.2 presents a case study of three scenarios with different data inputs.

3.4.1 Forecasting architecture

In the proposed approach, the SR measured by the Cranfield weather station is used as the reference weather parameter to select predictors from both weather stations. The selection of SR as the reference was made since it is the most significant predictor of PV generation. In step 1, the PCC function is used to calculate the correlation matrix of weather parameters reported by the on-site weather station in Cranfield. The weather parameters with moderate to strong positive correlations to SR are then selected as predictors for the forecasting model. By being strongly correlated to SR, this implies that these weather parameters will also have a strong linear relationship with PV generation and therefore are good predictors. In step 2, the correlation matrix is determined for the recorded weather parameters from the Bedford station. Again, the weather parameters with moderate to strong positive correlations to SR are selected as predictors. In step 3, the predictors from steps 1 and 2 are combined and along with the response, which is the historical PV generation supplied by the PV power plant meter, are selected as inputs into the RLA. However, only data points with the corresponding hourly time stamps across the three data sources are selected. This means that the resulting sampling rates of both the input and output variables are synchronized to 1 sample per hour. Furthermore, the system delay (or system lag) between input and output variables are mainly caused by the algorithm execution time (or required computation time) which is insignificant in comparison to the sample interval (one hour). The input data is then pre-processed using Equation (3-2) before being separated, using 75% for training and the remaining 25% is for model testing. The RLA flowchart presented in Figure 3-1, trains and test different forecasting algorithms and identifies the best

forecasting model based on the lowest RMSE in the test sample. Figure 3-2 shows an illustration of the forecasting architecture.

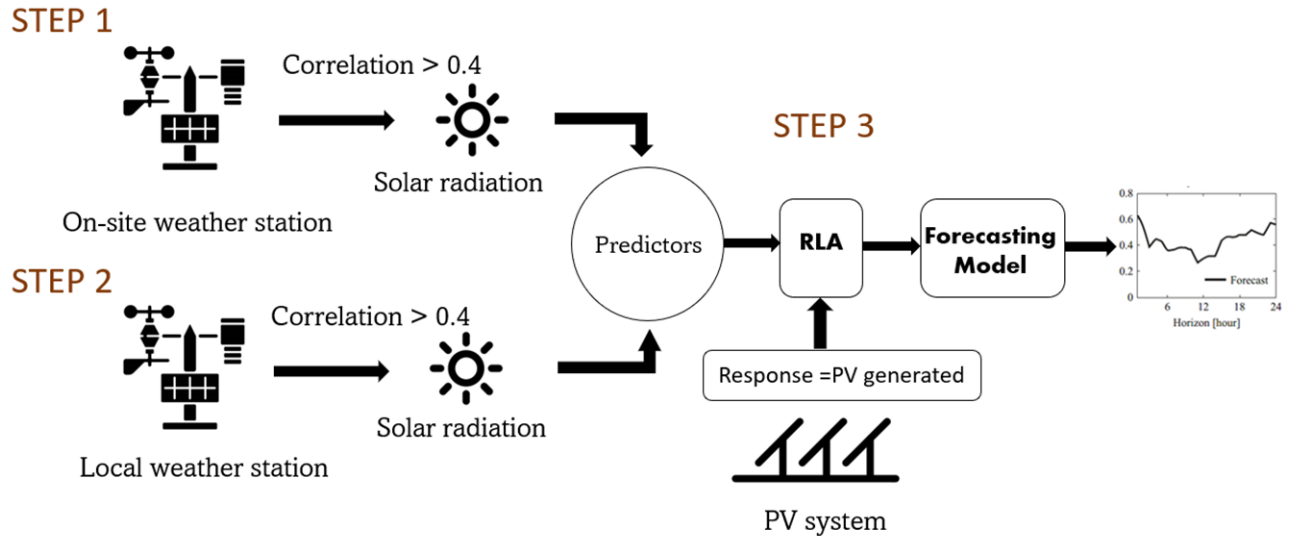


Figure 3-2 Proposed three-step forecasting architecture

3.4.2 Case study description

Three scenarios are examined, and the outcomes are compared, to gauge the effectiveness of the forecasting architecture.

- a. Scenario 1: Cranfield correlated weather data.
- b. Scenario 2: Bedford correlated weather data.
- c. Scenario 3: Cranfield and Bedford correlated weather data.

The above scenarios will investigate whether using LLDF to combine correlated data from the two stations increases the forecasting model's accuracy and therefore is a valid component of the proposed forecasting architecture.

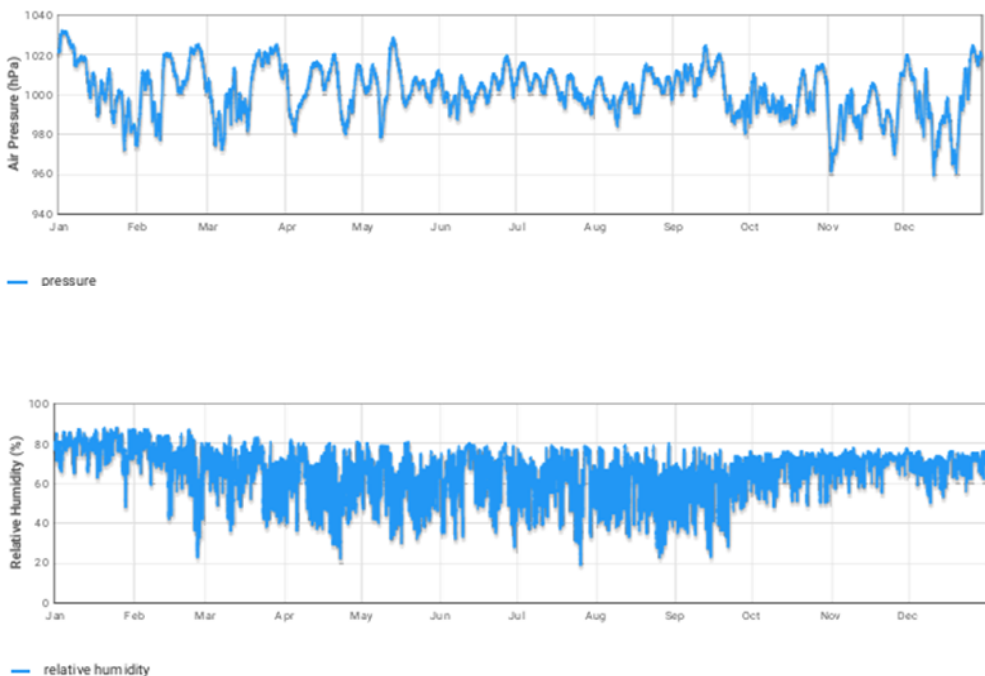
3.5 Results and discussion

The case study and methods given in the previous sections are summarized here. The outcome of the correlation matrix of the on-site meteorological variables is

shown in Section 3.5.1. The correlation matrix of the weather variables captured by the Bedford weather station is shown in Section 3.5.2. Based on RMSE and R^2 values for the three case scenarios mentioned above, Section 3.5.3 shows which prediction model is the most accurate. Additionally, a prediction for the following day is used to demonstrate how well the model performs.

3.5.1 Cranfield correlation matrix

According to the first step of the proposed forecasting architecture, weather parameters from the on-site weather station that have moderate to strong positive correlations to SR are selected based on coefficients above 0.4.



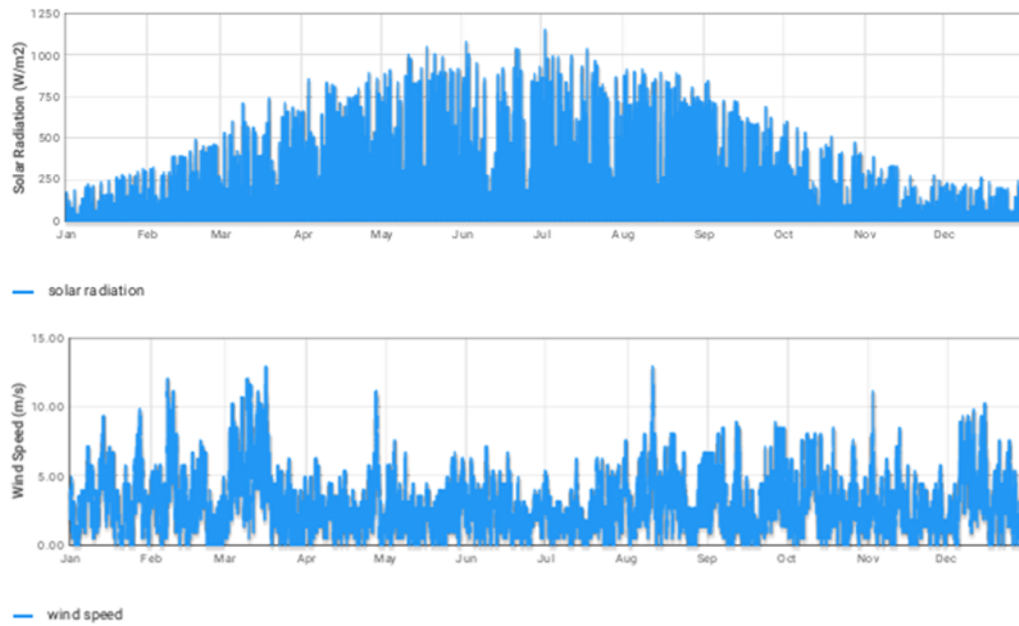
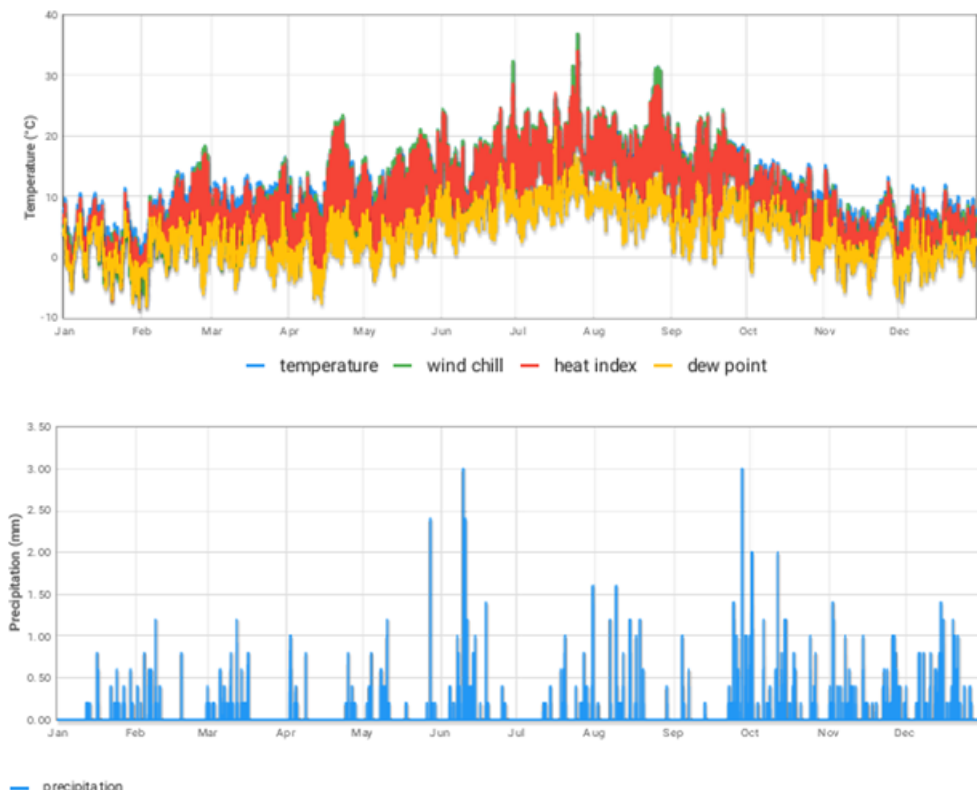


Figure 3-3 2019 Weather parameters from the Kisanhum Weather Station in Cranfield University shows a graphical representation of the weather parameters from the Kisanhub Weather Station in 2019.



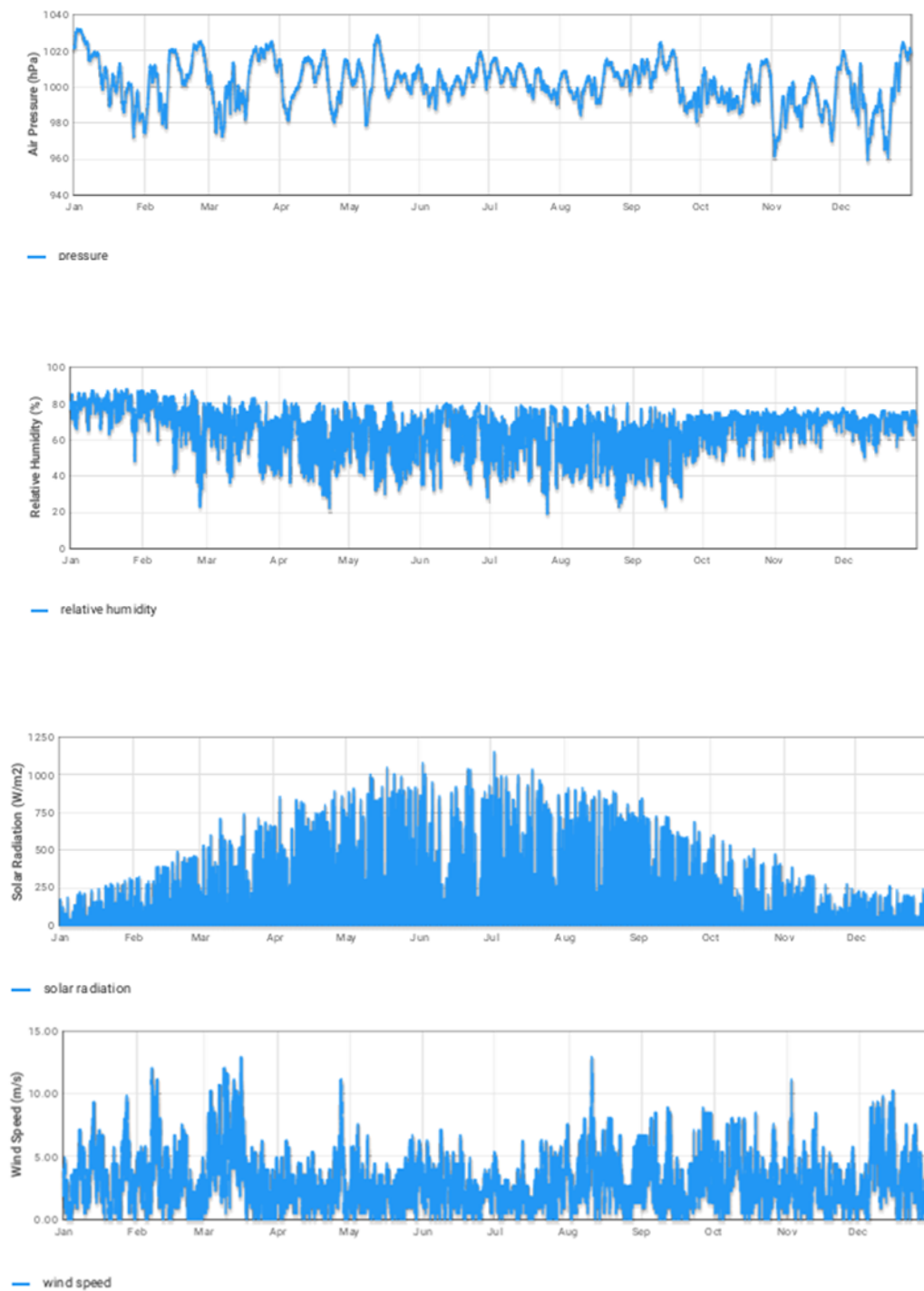


Figure 3-3 2019 Weather parameters from the Kisanhum Weather Station in Cranfield University [62]

Using the PCC function, the correlation matrix is obtained. As can be seen in Table 3-4, three weather parameters, Heat Index, Temperature and Wind Chill, are selected based on coefficients above 0.4.

Table 3-4 Correlation correlation matrix

	<i>Solar Radiation</i>
Solar Radiation	1
Wind Speed	0.1512
Relative Humidity	-0.5910
Pressure	0.1483
Precipitation	-0.0677
Dew Point	0.2164
Heat Index	0.5039
Temperature	0.5184
Wind Chill	0.4993

This step is crucial in finding the predictors (SR included) or input variables into the forecasting algorithm for Scenario 1 and one-half of the predictors for Scenario 3. It is not surprising to see that the above weather parameters were selected. SR has a strong linear relationship to outdoor surface air temperature [78,79]. Heat index is also called apparent temperature [80] and wind chill can also be expressed as wind chill temperature; a measure of human sensation to cold weather temperature [81]. All three weather parameters are related to the measurement of outdoor temperature.

3.5.2 Bedford correlation matrix

Similarly, from the Bedford station, the PCC function is used to obtain the correlation matrix where the Hourly Temperature, Hourly Wet Bulb Temperature and Hourly Global Radiation all had moderate to strong positive coefficients to SR according to section 3.3.2. Results are displayed in Table 3-5.

Table 3-5 Correlation matrix of weather parameters in Bedford

	<i>Solar Radiation</i>
Solar Radiation	1
Hourly Pressure at Mean Sea Level	0.1237
Hourly Dewpoint Temperature	0.1796
Hourly Rainfall Total	-0.0567
Hourly Relative Humidity	-0.6463
Hourly Mean Wind Direction	-0.0470
Hourly Mean Windspeed	0.1275
Hourly Maximum Gust	0.1688
Hourly Temperature	0.5383

Hourly Wet Bulb Temperature	0.4067
Hourly Global Radiation	0.9367
Hourly Total Cloud Cover	-0.1631
Hourly Visibility	0.1779

In this step, the input variables for Scenario 2 and the other half of Scenario 3 are identified. Parameters related to temperature are once again selected. In a similar assessment of twenty-four weather parameters, Fraihat et al. [82] used a correlation coefficient benchmark of 0.5 to identify which parameter had a strong correlation to SR. Temperature and radiation were among eight out of the twenty-four parameters that fit this criterion. This is consistent with the analysis here.

Comparing Table 3-4 and Table 3-5, it is easily deduced that the coefficients for SR, Pressure, Dew point, Relative Humidity, Wind Speed, and temperature are approximately the same across both stations. For instance, Relative Humidity for both stations are approximately -0.6 while Temperature are approximately 0.5. This is important because if the coefficients were considerably different given the relative proximity of both weather stations to each other, the data would have to be queried. Therefore, another function of PCC is to cross-validate data to ensure accuracy and reliability.

3.5.3 Performance of forecasting architecture

The weather parameters highlighted in Section 3.5.1 and Section 3.5.2 are the predictors that become the input into the RLA. Figure 3-4 shows the block diagram of the three-step forecasting architecture with the results of section 3.5.1 and 3.5.2. Table 3-6 shows a summary of the training input parameters for each scenario when using the RLA, with the historical PV energy selected as the response for all of them.

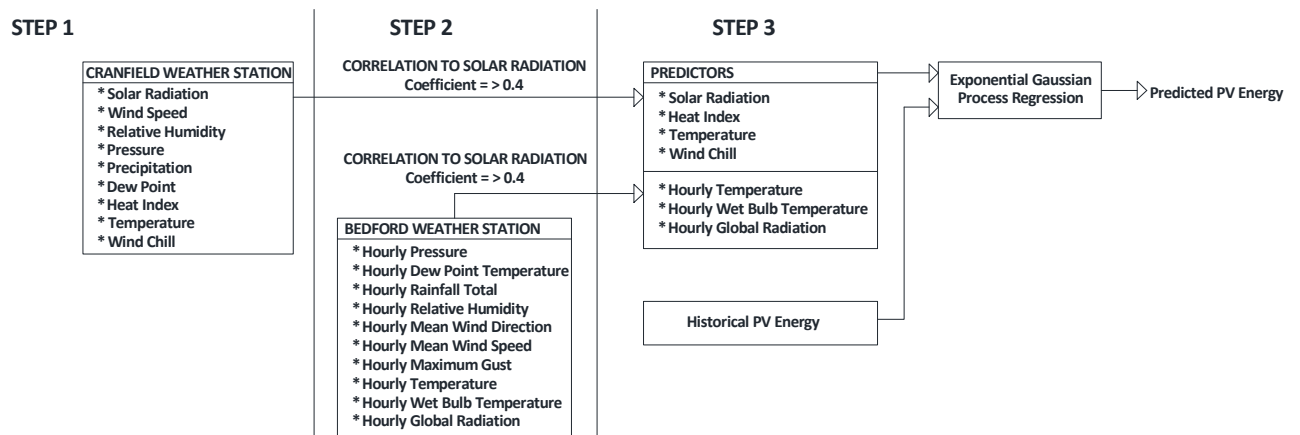


Figure 3-4 Case study: Three-step forecasting architecture block diagram

Table 3-6 RLA model training input parameters

Training Input	Scenario 1	Scenario 2	Scenario 3
Predictors	*Solar Radiation *Heat Index *Temperature *Wind Chill	*Hourly Temperature *Hourly Wet bulb Temperature *Hourly Global Radiation	*Solar Radiation *Heat Index *Temperature *Wind Chill *Hourly Temperature *Hourly Wet bulb Temperature *Hourly Global Radiation
Response	PV energy generated (all scenarios)		

According to Table 3-6, Scenario 1 has four predictors, Scenario 2 has three, while Scenario 3, which combines predictors from Scenarios 1 and 2 has seven. Each scenario is analyzed separately using the RLA flowchart in **Error! Reference source not found.** The RLA trains over twenty forecasting algorithms and identifies the best performing based on the lowest RMSE. To give an example of this, Scenario 3 is analyzed, and the results are shown in Table 3-7 (see Table A-1 and Table A-2 in Appendix A for similar analysis on Scenarios 1 and 2 respectively).

Table 3-7 Performance comparison of regression models in Regression learner app for Scenario 3

	Train				Test			
Model	MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²

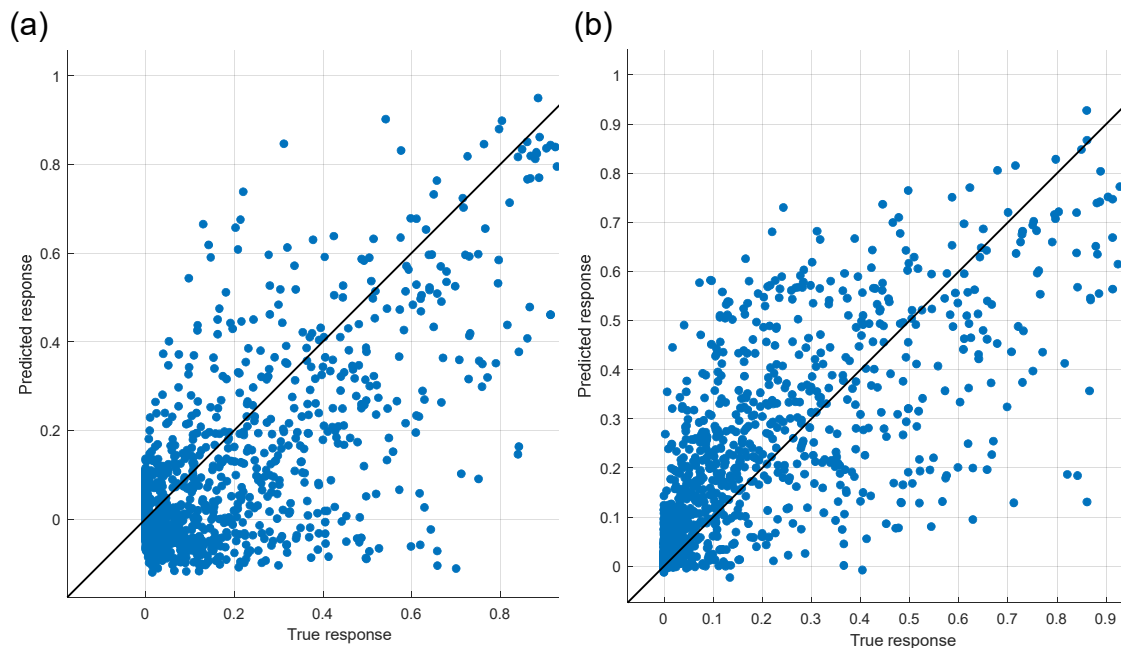
LINEAR REGRESSION								
Linear	0.0724	0.0124	0.1114	0.81	0.0975	0.0138	0.1176	0.57
Interactions Linear	0.0645	0.0109	0.1044	0.83	0.0904	0.0167	0.1293	0.49
Robust Linear	0.0741	0.0191	0.1381	0.71	0.0489	0.0090	0.0950	0.72
Stepwise Linear	0.0646	0.0109	0.1045	0.83	0.0889	0.0162	0.1273	0.50
REGRESSION TREES								
Fine Tree	0.0339	0.0044	0.0664	0.93	0.0495	0.0116	0.1079	0.64
Medium Tree	0.0474	0.0081	0.0901	0.88	0.0451	0.0098	0.0992	0.70
Coarse Tree	0.0560	0.0104	0.1019	0.84	0.0432	0.0087	0.0932	0.73
SUPPORT VECTOR MACHINES								
Linear SVM	0.0676	0.0129	0.1135	0.80	0.0707	0.0097	0.0985	0.70
Quadratic SVM	0.0598	0.0111	0.1054	0.83	0.0871	0.0178	0.1335	0.45
Cubic SVM	0.0573	0.0010	0.0999	0.85	0.0914	0.0192	0.1384	0.41
Fine Gaussian	0.0433	0.0068	0.0828	0.90	0.0914	0.0137	0.1171	0.58
Medium Gaussian	0.0561	0.0098	0.0992	0.85	0.0608	0.0102	0.1008	0.69
Coarse Gaussian	0.0644	0.0119	0.1092	0.82	0.0504	0.0074	0.0862	0.77
GAUSSIAN PROCESS REGRESSION								
Rational Quadratic	0.0007	3.05e-06	0.0017	1.00	0.0948	0.0213	0.1459	0.34
Squared Exponential	0.0013	8.71e-06	0.0030	1.00	0.1411	0.0447	0.2115	-0.38
Matern 5/2	0.0004	9.30e-07	0.0010	1.00	0.0966	0.2194	0.1481	0.33
Exponential	5.41e-05	1.20e-08	0.0001	1.00	0.0639	0.0111	0.1054	0.66
ENSEMBLES OF TREES								
Boosted Trees	0.0571	0.0109	0.1042	0.82	0.0448	0.0074	0.0859	0.84
Bagged Trees	0.0413	0.0059	0.0765	0.91	0.0436	0.0082	0.0906	0.75
NEURAL NETWORK								
Narrow Neural	0.0561	0.0010	0.0995	0.85	0.0803	0.0146	0.1209	0.55
Medium Neural	0.0535	0.0086	0.0928	0.87	0.1109	0.0206	0.1436	0.37
Wide Neural	0.0491	0.0071	0.0841	0.89	0.1284	0.0295	0.1718	0.09
Bilayered Neural	0.0528	0.0091	0.0956	0.86	0.0631	0.0107	0.1036	0.67
Trilayered Neural	0.0491	0.0081	0.0899	0.88	0.0940	0.0235	0.1534	0.28

Table 3-7 shows that Boosted Trees had the lowest RMSE values for the testing phase. This model is then used to analyze Scenarios' 1 and 2 and the results are compared to the result in Table 3-7 and are displayed in Table 3-8.

Table 3-8 Performance comparison of case study scenarios using Boosted Trees

Scenario	Test				Train			
	MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²
Scenario 1	0.0615	0.0118	0.1087	0.82	0.0441	0.0084	0.0917	0.74
Scenario 2	0.0663	0.0128	0.1133	0.80	0.0465	0.0097	0.0987	0.70
Scenario 3	0.0571	0.0109	0.1042	0.82	0.0448	0.0074	0.0859	0.84

According to Table 3-8, Scenario 3 had the lowest RMSE values for both training (0.1042) and testing (0.0859). This proves that by combining strong predictors from more than one source through LLDF, the accuracy of the forecast is improved. This is likely because forecasting depends on the number of relevant inputs in the forecasting architecture that correlate strongly with the forecasting output. Figure 3-5 shows a graphical illustration of the data points between the predicted and the actual responses for the three scenarios using the Boosted Tree prediction algorithm. The trendline of Scenario 3 is more accurate than the other scenarios because its R-square value of 0.84 is the closest to 1. Further confirmation of this is how the data points are packed closer to the trendline in Scenarios 3 than the other scenarios.



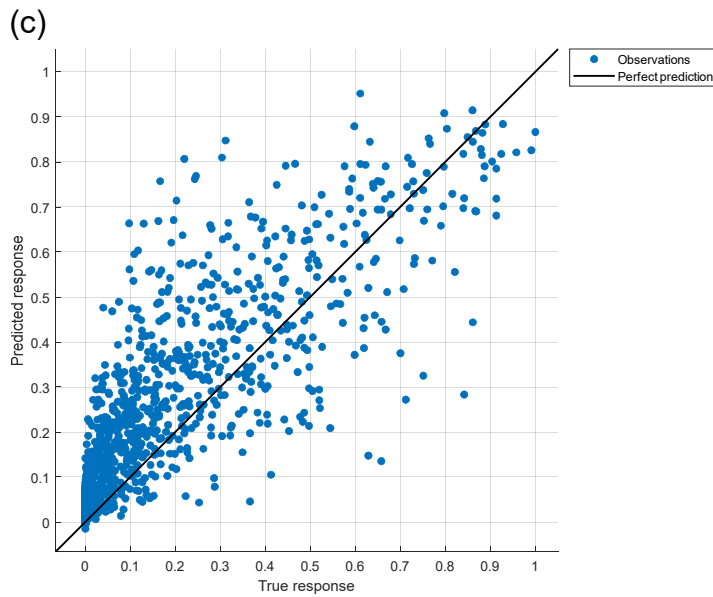
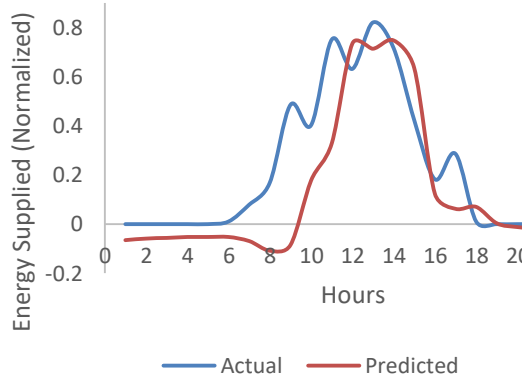


Figure 3-5 Model performance of Boosted Trees ((a) Scenario 1 (b) Scenario 2 (c) Scenario 3)

Table 3-8 and Figure 3-5 shows the model's performance in learning how to predict the PV generation. To investigate its forecasting ability further, we evaluate the prediction of a randomly selected day and determine how close it is to the actual PV generated. Figure 6 presents a comparison of the actual and the predicted amount of energy supplied from the PV under Scenarios 1 to 3. It is clear to see that Scenario 3 (Figure 3-6c) provides the best approximation of the actual PV generated than any of the other 2 scenarios. This is because the PV energy supplied in Scenario 1 (Figure 3-6a) appears to be underestimated between 0 and 12 hrs, and the afternoon peak between 16-18 hrs is also not represented. In Scenario 2 (Figure 3-6b), the night and morning energy supply (0-12 hrs) are relatively well represented. However, there is a substantial dip in the amount of energy predicted between 12 and 13 hrs. There is evidence that the model predicts small energy supply peak between 16-18 hrs. In Scenario 3, the prediction for the night and morning energy supply are in good agreement with the actual data. There is a small dip in the forecasted energy supply between 12 and 13 hrs, however, it is of a much smaller magnitude than that observed in Scenario 2. The afternoon peak in energy supply is small.

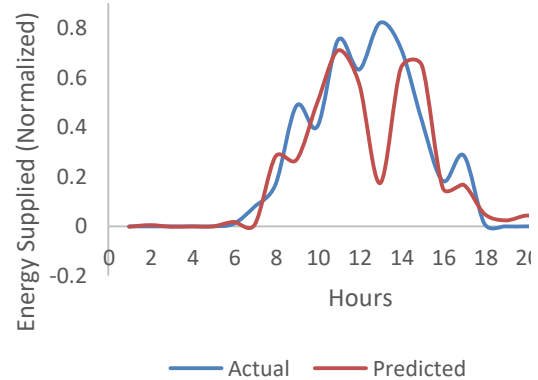
(a)

Scenario 1 : 12th September 20



(b)

Scenario 2 : 12th September 20



(c)

Scenario 3 : 12th September 2019

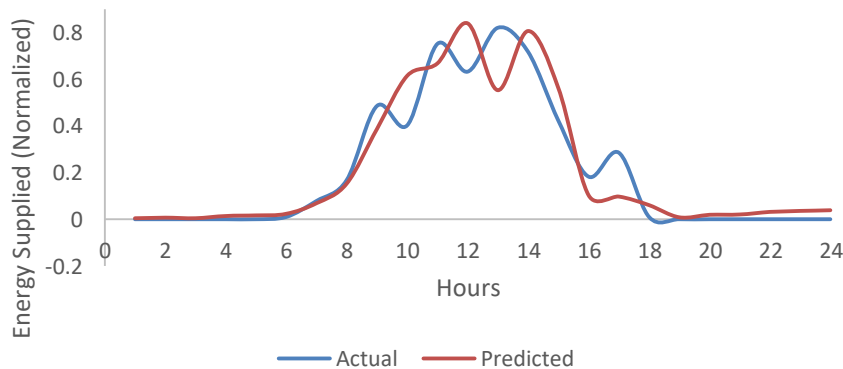


Figure 3-6 Actual and predicted PV generation on 12th September 2019 ((a) Scenario 1 (b) Scenario 2 (c) Scenario 3)

Beyond visually comparing the performance of the scenarios in Figure 3-6, further insight can be provided by estimating the area under the predicted curve of each scenario and comparing it to the area under the actual curve. The scenario that has the closest approximation of its predicted curve to the actual curve has the better performance and accuracy. This area is simply the amount of energy supplied (normalized) per day. For a simple approximation of this area, the Trapezoidal Rule is used. This rule approximates the area under the graph by dividing it into smaller trapezoids. The total area under the curve is equal to the

sum of the area of the trapezoids under the graph. The Trapezoidal Rule is shown in Equation (3-7) [83].

$$\int_a^b f(x)dx \approx \left(\frac{b-a}{2}\right)(f(a) + f(b)) \quad (3-7)$$

In Equation (3-7), limits a and b refer to the hours on the x axis in Figure 3-6. This represents the width of the trapezoid. Values $f(a)$ and $f(b)$ represent the corresponding energy supplied (normalized) at hour ‘ a ’ and hour ‘ b ’ respectively. Equation (3-7) was applied to the three scenarios (See Table B-1 in Appendix A for calculations) to compare their approximation of the actual energy supplied. The results show that the area under the curve of the actual energy supplied is approximately 4.969. Ranking the scenario’s based on the closest approximation to the actual area, Scenario 3 is closest with an area of 5.115, followed by Scenario 2 with 4.424 and Scenario 1 with 2.906. This result and the order of performance validates the test result in Table 3-8 in which Scenario 3 (0.0859) outperforms the other scenarios, and Scenario 2 (0.0987) outperforms Scenario 1 (0.0917).

A further proof of the versatility of the proposed forecasting architecture can be made with a longer forecast. Section 3.1 references the different types of forecasting time horizons. Figure 3-7 shows how the proposed forecasting architecture of Scenario 3 is applied to a day-ahead short-term forecast. Here, a random week in 2019 is selected to show the performance of the forecasting architecture when applied to a medium-term forecast. The graph shows a good enough approximation, which means that the proposed architecture is flexible and can be used for predicting longer time horizons.

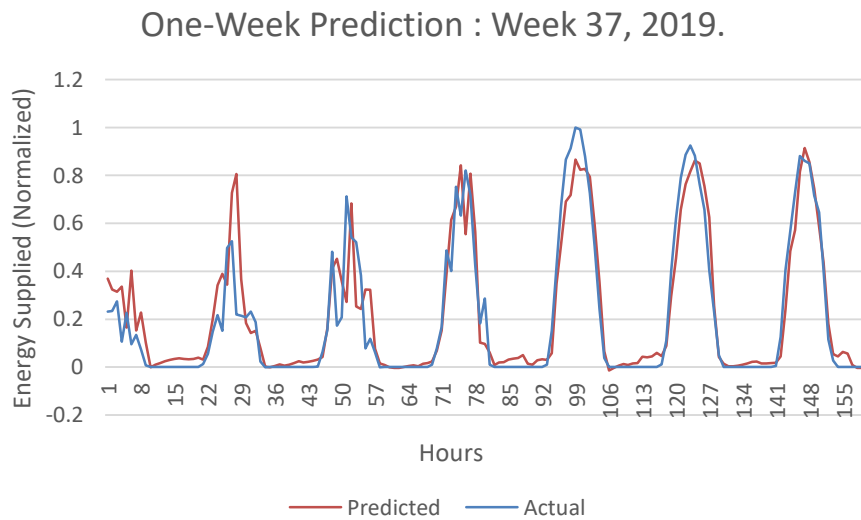


Figure 3-7 Actual and predicted PV generation for Scenario 3 on the 37th week in 2019.

3.5.4 Validation of forecasting architecture

To validate the model, we apply the forecasting architecture in predicting PV generation of the three rooftop solar arrays described in Section 3.3.1. The results are displayed in Table 3-9.

Table 3-9 Validation results of proposed architecture using residential solar arrays

BUILDINGS	MODEL	RMSE	MSE	MAE	R ²
B146	Boosted Trees	0.0885	0.0078	0.0499	0.68
B242	MODEL	RMSE	MSE	MAE	R ²
	Boosted Trees	0.0984	0.0097	0.0547	0.67
B052A	MODEL	RMSE	MSE	MAE	R ²
	Boosted Trees	0.1425	0.0203	0.0811	0.52

In Table 3-9, the results show that the proposed architecture maintains its accuracy in predicting PV generation as evident by the low RMSE values. Another important fact is that the proposed architecture made the predictions

across three different PV system sizes and data time periods. By confirming its accuracy for both commercial and residential applications, the models can be universally adopted for different DDG clients using different historical weather data as inputs into the model.

3.6 Conclusion

The proposed three-step forecasting architecture that uses low-level data fusion to combine weather parameters from both an on-site and a nearby local weather station outperforms models using a singular data source. Pearson Correlation Coefficient was used to identify weather variables with moderate to strong positive correlations to solar radiation. Combining useful data from the two weather stations increased the number of strong predictors in the forecasting model and improved its performance. In addition, raw meteorological data was pre-processed using the Rescaling and List-wise deletion methods, which proved to be effective in increasing the accuracy of the model. The proposed architecture is tested on a utility scale solar plant (1 MW), showing a 6% and 13% prediction accuracy improvement when compared to solely using data from the on-site and local weather station respectively. It is further validated using data from three residential rooftop solar systems (8 kW, 10.5 kW and 15 kW), achieving root-mean square values of 0.0859, 0.0984, 0.1425 and 0.0885 respectively. The low root-mean square values, a measure of the predicted PV to actual PV generation, proves that the model can be adopted for different PV plant sizes.

The limitation of the forecasting architecture is that it only works if the local weather station and the on-site PV plant measure the reference weather parameter, which in this case is solar radiation. In the absence of this, an alternate reference parameter with strong correlation to PV generation needs to be determined. Additionally, the scope of the study was limited to data within the UK. To test the robustness of the forecasting architecture, other PV plant in different locations outside the UK could be considered. In addition, only weather parameters with strong positive correlations to the reference parameter were considered. In a future study, the impact of weather parameters with strong

negative correlations should also be analyzed. Finally, the optimization of the RLA methodology can be investigated by exploring the impact of other validation methods like Cross-validation and Holdout validation to see if they directly improve the accuracy of the prediction model.

Credit author statement

Tolulope O. Falope: Conceptualization, Methodology, Software, Data curation, Formal Analysis, Visualization, Writing – original draft, **Lao Liyun:** Writing – review & editing, Supervision, **Dawid Hanak:** Writing – review & editing, Supervision.

References

- [1] Banik R, Biswas A. Improving Solar PV Prediction Performance with RF-CatBoost Ensemble: A Robust and Complementary Approach. *Renewable Energy Focus* 2023;46:207–21. <https://doi.org/10.1016/J.REF.2023.06.009>.
- [2] Abisoye BO, Sun Y, Zenghui W. A survey of artificial intelligence methods for renewable energy forecasting: Methodologies and insights. *Renewable Energy Focus* 2024;48:100529. <https://doi.org/10.1016/J.REF.2023.100529>.
- [3] Katz J, Cochran J. Integrating Variable Renewable Energy into the Grid: Key Issues Greening the Grid. National Renewable Energy Laboratory 2015;NREL/FS-6A20-63033.
- [4] Basmadjian R, Niedermeier F, de Meer H. Demand-Side Flexibility and Supply-Side Management: The Use Case of Data Centers and Energy Utilities 2017:187–204. https://doi.org/10.1007/978-3-319-65082-1_9.
- [5] Cox S, Xu K. Integration of Large-Scale Renewable Energy in the Bulk Power System: Good Practices from International Experiences. National Renewable Energy Laboratory 2020.

- [6] Li B, Zhang J. A review on the integration of probabilistic solar forecasting in power systems. *Solar Energy* 2020;210:68–86. <https://doi.org/10.1016/J.SOLENER.2020.07.066>.
- [7] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews* 2018;81:912–28. <https://doi.org/10.1016/J.RSER.2017.08.017>.
- [8] Tian T, Chernyakhovskiy I. Forecasting Wind and Solar Generation: Improving System Operations, Greening the Grid. National Renewable Energy Laboratory 2016.
- [9] Meenal R, Binu · D, Ramya · K C, Prawin ·, Michael A, Vinoth Kumar · K, et al. Weather Forecasting for Renewable Energy System: A Review. *Archives of Computational Methods in Engineering* 2021;1:3. <https://doi.org/10.1007/s11831-021-09695-3>.
- [10] Gupta P, Singh R. International Journal of Sustainable Engineering ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/tsue20> PV power forecasting based on data-driven models: a review Priya Gupta & Rhythm Singh PV power forecasting based on data-driven models: a review 2021. <https://doi.org/10.1080/19397038.2021.1986590>.
- [11] Ma J, Ma X. A review of forecasting algorithms and energy management strategies for microgrids. *Systems Science & Control Engineering: An Open Access Journal* 2018;6:237–48. <https://doi.org/10.1080/21642583.2018.1480979>.
- [12] Carrera B, Sim M-K, Jung J-Y. PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data; PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data. *The Institution of Engineering and Technology* 2020;14:2192–201. <https://doi.org/10.1049/iet-rpg.2018.6174>.

- [13] Minhas DM, Khalid RR, Frey G. Load control for supply-demand balancing under Renewable Energy forecasting. 2017 IEEE 2nd International Conference on Direct Current Microgrids, ICDCM 2017 2017:365–70. <https://doi.org/10.1109/ICDCM.2017.8001071>.
- [14] Aghajani GR, Shayanfar HA, Shayeghi H. Demand side management in a smart micro-grid in the presence of renewable generation and demand response. Energy 2017;126:622–37. <https://doi.org/10.1016/J.ENERGY.2017.03.051>.
- [15] Manjarres D, Alonso R, Gil-Lopez S, Landa-Torres I. Solar Energy Forecasting and Optimization System for Efficient Renewable Energy Integration. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2017;10691 LNAI:1–12. https://doi.org/10.1007/978-3-319-71643-5_1.
- [16] Pawar P, TarunKumar M, Vittal K. P. An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. Measurement 2020;152:107187. <https://doi.org/10.1016/J.MEASUREMENT.2019.107187>.
- [17] Sangrody H, Zhou N, Zhang Z. Similarity-Based Models for Day-Ahead Solar PV Generation Forecasting. IEEE Access 2020;8:104469–78. <https://doi.org/10.1109/ACCESS.2020.2999903>.
- [18] Pawar P, Mithulananthan N, Raza MQ. Solar PV Power Forecasting Using Modified SVR with Gauss-Newton Method. Proceedings - 2020 IEEE 2nd Global Power, Energy and Communication Conference, GPECOM 2020 2020:226–31. <https://doi.org/10.1109/GPECOM49333.2020.9247935>.
- [19] Feng H, Yu C. A novel hybrid model for short-term prediction of PV power based on KS-CEEMDAN-SE-LSTM. Renewable Energy Focus 2023;47:100497. <https://doi.org/10.1016/J.REF.2023.100497>.
- [20] Wang B, Che J, Wang B, Feng S. A Solar Power Prediction Using Support Vector Machines Based on Multi-source Data Fusion. 2018 International

Conference on Power System Technology, POWERCON 2018 - Proceedings 2019:4573–7. <https://doi.org/10.1109/POWERCON.2018.8601672>.

[21] Robert C, Jessep W, Sutton JJ, Hicks TM, Loeffen M, Farouk M, et al. Evaluating low- mid- and high-level fusion strategies for combining Raman and infrared spectroscopy for quality assessment of red meat. *Food Chem* 2021;361:130154. <https://doi.org/10.1016/J.FOODCHEM.2021.130154>.

[22] Rammelkamp K, Schröder S, Kubitza S, Vogt DS, Frohmann S, Hansen PB, et al. Low-level LIBS and Raman data fusion in the context of in situ Mars exploration. *Journal of Raman Spectroscopy* 2020;51:1682–701. <https://doi.org/10.1002/JRS.5615>.

[23] Kuo PH, Huang CJ. A Green Energy Application in Energy Management Systems by an Artificial Intelligence-Based Solar Radiation Forecasting Model. *Energies* 2018, Vol 11, Page 819 2018;11:819. <https://doi.org/10.3390/EN11040819>.

[24] Narvaez G, Giraldo LF, Bressan M, Pantoja A. Machine learning for site-adaptation and solar radiation forecasting. *Renew Energy* 2021;167:333–42. <https://doi.org/10.1016/J.RENENE.2020.11.089>.

[25] Kushwaha V, Pindoriya NM. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast | Elsevier Enhanced Reader. Elsevier Ltd 2019;140:124–39. <https://doi.org/10.1016/j.renene.2019.03.020>.

[26] Anaadumba R, Liu Q, Marah BD, Nakoty FM, Liu X, Zhang Y. A renewable energy forecasting and control approach to secured edge-level efficiency in a distributed micro-grid. *Cybersecurity* 2021;4. <https://doi.org/10.1186/s42400-020-00065-3>.

[27] Rana M, Rahman A. Multiple steps ahead solar photovoltaic power forecasting based on univariate machine learning models and data re-sampling. *Sustainable Energy, Grids and Networks* 2020;21. <https://doi.org/10.1016/J.SEGAN.2019.100286>.

- [28] Sadeghi D, Golshanfard A, Eslami S, Rahbar K, Kari R. Improving PV power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons. *Renewable Energy Focus* 2023;45:242–58. <https://doi.org/10.1016/J.REF.2023.04.010>.
- [29] Sarp S, Kuzlu M, Cali U, Guler O. An Interpretable Solar Photovoltaic Power Generation Forecasting Approach Using An Explainable Artificial Intelligence Tool. *IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2021, p. 1–5. <https://doi.org/10.1109/ISGT49243.2021.9372263>.
- [30] Aprillia H, Yang H-T, Huang C-M. Short-Term Photovoltaic Power Forecasting Using a Convolutional Neural Network-Salp Swarm Algorithm. *Energies (Basel)* 2020;13:1–20. <https://doi.org/10.3390/en13081879>.
- [31] Cannizzaro D, Aliberti A, Bottaccioli L, Macii E, Acquaviva A, Patti E. Solar radiation forecasting based on convolutional neural network and ensemble learning. *Expert Syst Appl* 2021;181. <https://doi.org/10.1016/j.eswa.2021.115167>.
- [32] Nguyen NQ, Bui LD, Doan BV, Sanseverino ER, Cara DD, Nguyen QD. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. *Electric Power Systems Research* 2021;199. <https://doi.org/10.1016/j.epsr.2021.107427>.
- [33] Pan C, Tan J, Feng D. Prediction intervals estimation of solar generation based on gated recurrent unit and kernel density estimation. *Neurocomputing* 2021;453:552–62. <https://doi.org/10.1016/j.neucom.2020.10.027>.
- [34] Ngoc-Lan Huynh A, Deo RC, Ali M, Abdulla S, Raj N. Novel short-term solar radiation hybrid model: Long short-term memory network integrated with robust local mean decomposition. *Appl Energy* 2021;298. <https://doi.org/10.1016/j.apenergy.2021.117193>.

- [35] Pedregal DJ, Trapero JR. Adjusted combination of moving averages: A forecasting system for medium-term solar irradiance. *Appl Energy* 2021;298. <https://doi.org/10.1016/j.apenergy.2021.117155>.
- [36] Ahmad T, Zhang D, Huang C. Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications. *Energy* 2021;231. <https://doi.org/10.1016/j.energy.2021.120911>.
- [37] Korkmaz D. SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting. *Appl Energy* 2021;300. <https://doi.org/10.1016/j.apenergy.2021.117410>.
- [38] Najibi F, Apostolopoulou D, Alonso E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *International Journal of Electrical Power & Energy Systems* 2021;130:106916. <https://doi.org/10.1016/J.IJEPES.2021.106916>.
- [39] Rodríguez F, Fleetwood A, Galarza A, Fontán L. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renew Energy* 2018;126:855–64. <https://doi.org/10.1016/J.RENENE.2018.03.070>.
- [40] Thukral MK. Solar Power Output Prediction Using Multilayered Feedforward Neural Network: A Case Study of Jaipur. 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC), 2020, p. 1–6. <https://doi.org/10.1109/iSSSC50941.2020.9358821>.
- [41] Munawar U, Wang Z. A Framework of Using Machine Learning Approaches for Short-Term Solar Power Forecasting. *Journal of Electrical Engineering & Technology* 2020;15:561–9. <https://doi.org/10.1007/s42835-020-00346-4>.
- [42] Yakoubi H, el Mghouchi Y, Abdou N, Hajou A, Khellouki A. Correlating clearness index with cloud cover and other meteorological parameters for

forecasting the global solar radiation over Morocco. *Optik (Stuttg)* 2021;242. <https://doi.org/10.1016/j.ijleo.2021.167145>.

[43] Zang H, Cheng L, Ding T, Cheung KW, Liang Z, Wei Z, et al. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Gener Transm Distrib* 2018;12:4557–67. <https://doi.org/10.1049/iet-gtd.2018.5847>.

[44] Chen B, Lin P, Lai Y, Cheng S, Chen Z, Wu L. Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets. *Electronics (Basel)* 2020;9. <https://doi.org/10.3390/electronics9020289>.

[45] Iyengar S, Sharma N, Irwin D, Shenoy P, Ramamritham K. SolarCast-A Cloud-based Black Box Solar Predictor for Smart Homes. *BuildSys '14: 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, 2014, p. 40–9. <https://doi.org/10.1145/2674061.2674071>.

[46] Andrade JR, Bessa RJ. Improving Renewable Energy Forecasting With a Grid of Numerical Weather Predictions. *IEEE Trans Sustain Energy* 2017;8th:1571–80. <https://doi.org/10.1109/TSTE.2017.2694340>.

[47] Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Math Comput Simul* 2017;131:88–100. <https://doi.org/10.1016/J.MATCOM.2015.05.010>.

[48] Persson C, Bacher P, Shiga T, Madsen H. Multi-site solar power forecasting using gradient boosted regression trees. *Solar Energy* 2017;150:423–36. <https://doi.org/10.1016/J.SOLENER.2017.04.066>.

[49] Abedinia O, Raisz D, Amjady N. IET Renewable Power Generation Effective prediction model for Hungarian small-scale solar power output. *The Institution of Engineering and Technology* 2017;11:1648–58. <https://doi.org/10.1049/iet-rpg.2017.0165>.

- [50] Kim S-G, Jung J-Y, Sim MK. A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning. *Sustainability* 2019. <https://doi.org/10.3390/su11051501>.
- [51] Kyliashkina IA, Eroshenko SA, Shelyug S. Intelligent Systems as a Tool for Predicting Electrical Energy and Power Generation. 2019 60th International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), 2019, p. 1–5. <https://doi.org/10.1109/ITMS47855.2019.8940721>.
- [52] Kazem HA, Yousif JH, Chaichan MT, Al-Waeli AHA, Sopian K. Long-term power forecasting using FRNN and PCA models for calculating output parameters in solar photovoltaic generation. *Heliyon* 2022;8:e08803. <https://doi.org/10.1016/J.HELIYON.2022.E08803>.
- [53] Chaudhary P, Rizwan M. Energy management supporting high penetration of solar photovoltaic generation for smart grid using solar forecasts and pumped hydro storage system. *Renew Energy* 2018;118:928–46. <https://doi.org/10.1016/J.RENENE.2017.10.113>.
- [54] Elma O, Taşcıkaraoğlu A, Tahir İnce A, Selamoğulları US. Implementation of a dynamic energy management system using real time pricing and local renewable energy generation forecasts. *Energy* 2017;134:206–20. <https://doi.org/10.1016/J.ENERGY.2017.06.011>.
- [55] Scabbia G, Sanfilippo A, Bachour D, Perez-Astudillo D. Exogenous Parameters in Solar Forecasting. *Conference Record of the IEEE Photovoltaic Specialists Conference* 2020;2020-June:0894–6. <https://doi.org/10.1109/PVSC45281.2020.9300800>.
- [56] Moreno-Muñoz A editor. *Large scale grid integration of renewable energy sources*. Institution of Engineering and Technology; 2017.
- [57] Gu B, Shen H, Lei X, Hu H, Liu X. Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method. *Appl Energy* 2021;299. <https://doi.org/10.1016/j.apenergy.2021.117291>.

- [58] Li LL, Wen SY, Tseng ML, Wang CS. Renewable energy prediction: A novel short-term prediction model of photovoltaic output power. *J Clean Prod* 2019;228:359–75. <https://doi.org/10.1016/J.JCLEPRO.2019.04.331>.
- [59] Wang J, Li P, Ran R, Che Y, Zhou Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Applied Sciences* 2018;8:1–14. <https://doi.org/10.3390/app8050689>.
- [60] Najibi F, Apostolopoulou D, Alonso E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *International Journal of Electrical Power and Energy Systems* 2021;130. <https://doi.org/10.1016/j.ijepes.2021.106916>.
- [61] de Jesus DAR, Mandal P, Velez-Reyes M, Chakraborty S, Senjyu T. Data Fusion Based Hybrid Deep Neural Network Method for Solar PV Power Forecasting. 51st North American Power Symposium, NAPS 2019 2019. <https://doi.org/10.1109/NAPS46351.2019.9000331>.
- [62] Cranfield University. Cranfield University's Living Laboratory and Urban Observatory - Data. Living Laboratory 2021. <https://www.livinglab.ac.uk/data.html> (accessed February 21, 2022).
- [63] Cranfield University. Energy Manager Live 2022. <https://cranfield.energymanagerlive.com/> (accessed February 21, 2022).
- [64] Microsoft. PEARSON function. Microsoft Support 2021. <https://support.microsoft.com/en-us/office/pearson-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18> (accessed October 15, 2021).
- [65] Microsoft. Use the Analysis ToolPak to perform complex data analysis - Office Support. Microsoft 2021. <https://support.microsoft.com/en-us/office/use-the-analysis-toolpak-to-perform-complex-data-analysis-6c67ccf0-f4a9-487c-8dec-bdb5a2cefab6> (accessed July 15, 2021).
- [66] Qin L, Huang W, Guo Z, Zhou Y, Zhang B. Topology Identification Method of Low-voltage Distribution Network Based on Improved Pearson Correlation

- | | | |
|-------------|--------|-----------|
| Coefficient | Method | 2022:1–6. |
|-------------|--------|-----------|
- <https://doi.org/10.1109/CIYCEE53554.2021.9676722>.
- [67] Qian K, Wang X, Yuan Y. Research on regional short-term power load forecasting model and case analysis. *Processes* 2021;9. <https://doi.org/10.3390/PR9091617>.
- [68] Herwanto HW, Handayani AN, Wibawa AP, Chandrika KL, Arai K. Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition. 7th International Conference on Electrical, Electronics and Information Engineering (ICEEE) 2021:1–3. <https://doi.org/10.1109/ICEEIE52663.2021.9616665>.
- [69] Normalize data - MATLAB normalize - MathWorks United Kingdom n.d. <https://uk.mathworks.com/help/matlab/ref/double.normalize.html> (accessed October 30, 2021).
- [70] Ghofrani M, Alolayan M. Time Series and Renewable Energy Forecasting. *Time Series Analysis and Applications* 2017. <https://doi.org/10.5772/INTECHOPEN.70845>.
- [71] Zazoum B. Solar photovoltaic power prediction using different machine learning methods. *Energy Reports* 2022;8:19–25. <https://doi.org/10.1016/J.EGYR.2021.11.183>.
- [72] Ekici S, Unal F, Ozleyen U. Comparison of different regression models to estimate fault location on hybrid power systems. *IET Generation, Transmission and Distribution* 2019;13:4756–65. <https://doi.org/10.1049/IET-GTD.2018.6213>.
- [73] Sunori SK, Negi PB, Mittal A, Arora S, Sharma S, Maurya S, et al. Prediction of pH of Soil Using Linear Regression, Decision Tree, GPR, and Gaussian SVM. *IEMECON 2021 - 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks* 2021. <https://doi.org/10.1109/IEMECON53809.2021.9689193>.

- [74] Ortigoza G, Zapata U. Covid-19 Projections: A Simple Machine Learning Approach. 2021 IEEE International Conference on Engineering Veracruz, ICEV 2021 2021. <https://doi.org/10.1109/ICEV52951.2021.9632647>.
- [75] Madhukumar M, Sebastian A, Liang X, Jamil M, Shabbir MNSK. Regression Model-Based Short-Term Load Forecasting for University Campus Load. IEEE Access 2022;10:8891–905. <https://doi.org/10.1109/ACCESS.2022.3144206>.
- [76] Popolo PS, Johnson AM. Relating Cepstral Peak Prominence to Cyclical Parameters of Vocal Fold Vibration from High-Speed Videoendoscopy Using Machine Learning: A Pilot Study. Journal of Voice 2021;35:703–16. <https://doi.org/10.1016/J.JVOICE.2020.01.026>.
- [77] dela Cruz MAA, Legaspi KMC, Marcelino RMD, Rosete JR, Sangalang DAR, Suarez CG, et al. Joint Gait Kinematic and Kinetic Analysis using Inertial Measurement Units and Plantar Pressure Sensor System. 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, HNICEM 2019 2019. <https://doi.org/10.1109/HNICEM48295.2019.9072701>.
- [78] Daut I, Yusoff MI, Ibrahim S, Irwanto M, Nsurface G. Relationship between the Solar Radiation and Surface Temperature in Perlis. Adv Mat Res, Durnten-Zurich: Trans Tech Publications; 2012, p. 143–7.
- [79] Shang K, Liu Y, Wang D, Li T. Dynamic relationship between outdoor air temperature and solar radiation. Tumu Jianzhu Yu Huanjing Gongcheng/Journal of Civil, Architectural and Environmental Engineering 2015;37:116–21. <https://doi.org/10.11835/J.ISSN.1674-4764.2015.05.017>.
- [80] Lateef KF, Paul J, Jahan Z. Weather Parameters and Heat Index Prediction Applying Deep Neural Networks. TENSYP 2021 - 2021 IEEE Region 10 Symposium 2021. <https://doi.org/10.1109/TENSYP52854.2021.9550852>.

- [81] Howarth ME, Laird NF. Intraseasonal variations of winter wind chill temperatures across Canada and the United States. *J Appl Meteorol Climatol* 2017;56:2951–62. <https://doi.org/10.1175/JAMC-D-16-0394.1>.
- [82] Fraihat H, Almbaideen AA, Al-Odienat A, Al-Naami B, de Fazio R, Visconti P. Solar Radiation Forecasting by Pearson Correlation Using LSTM Neural Network and ANFIS Method: Application in the West-Central Jordan. *Future Internet* 2022, Vol 14, Page 79 2022;14:79. <https://doi.org/10.3390/FI14030079>.
- [83] Altwallbah NMM, Radzi MAM, Azis N, Shafie S, Zainuri MAAM. New perturb and observe algorithm based on trapezoidal rule: Uniform and partial shading conditions. *Energy Convers Manag* 2022;264:115738. <https://doi.org/10.1016/J.ENCONMAN.2022.115738>.

4 Data preprocessing techniques for statistical PV forecasting models³

ABSTRACT

Data preprocessing is a critical step in improving the training and prediction performance of solar forecasting models. While data preprocessing is a widely researched topic, a structured scope for data-driven statistical PV forecasting models is lacking. This study surveys and classifies various data preprocessing techniques into categories relevant to statistical PV forecasting methods thus providing a structured preprocessing framework. The study also investigates the impact of these techniques on raw meteorological data obtained from two weather stations and PV generation data obtained from a solar farm. By providing an empirical analysis on an existing statistical forecasting model and a base-case scenario, the effects of combining various data preprocessing techniques are assessed. The results demonstrate that by combining data preprocessing methods, the accuracy of the existing model that uses rescaling normalization, list-wise deletion method, Pearson correlation coefficient, and low-data level fusion data preprocessing techniques is increased by 75% while the base-case scenario that uses either list-wise deletion method or linear interpolation to remove missing data is increased by 100%.

Keywords: data preprocessing, renewable energy sources, statistical PV method, machine learning, PV forecasting

³ This chapter is a paper authored by Tolulope Falope, Liyun Lao, & Da Huo, submitted to the Solar Energy journal.

4.1 Introduction

Data-driven models rely heavily on large amounts of data to improve accuracy. This implies that the quality of data is critical and needs to be considered. According to Das et al. [1], “the accuracy of the forecasting model can be considerably improved by the preprocessing of the input data”. But raw data is not always clean and may contain incompatible, redundant, unstructured, and/or missing data. Data preprocessing (DP) is primarily used to remove unnecessary information, simplify data as much as possible, increase the detectability of characteristics associated to the target response, and boost the accuracy of feature extraction and recognition [2]. DP takes raw data and converts it into meaningful and useful information for the prediction algorithm [3]. DP methods have been demonstrated to significantly improve model efficiency and accuracy when combined [4]. The question of what DP method(s) is always relevant to forecasting. Some researchers believe that selection of DP methods is dependent on the characteristics of the dataset [5–7]. Others believe to achieve an optimal outcome, experimental studies must be carried out to determine the overall arrangement and combination of DP techniques [8].

Several studies have explored the impact of DP methods on PV forecasting. Ranjan et al. [9] focuses on outlier detection and correction methods for time-series data. The authors identify and categorize these outlier methods into statistical-based, density-based, and distance-based methods. Bramm et al. [10] investigates outlier detection and removal, daylight-free selection, removal of linearly correlated dependencies, and handling of missing data. A comparison of a linear regression method before and after preprocessing showed a root mean square error (RMSE) reduction by 75.9%. In another study, Phan et al. [11] applied normalization, scaling, data filling, extreme gradient boosting (XGBoost), feature engineering, and Kernel Principal Component Analysis (KPCA) to a gated recurrent unit network (GRU). The proposed model with DP methods performed better than the deep learning GRU without pre and post processing. This study identifies sixteen DP models to be analysed which are: (i) List-wise deletion method (LWDM), (ii) Linear interpolation method (LIM), (iii) Delete outliers (DO), (iv) Fill outliers (FO), (v) Data smoothing (DS), (vi) Z-score (ZS), (vii) Rescaling method (RM), (viii) Low-level data fusion (LLDF), (ix) Pearsons Correlation Coefficient (PCC), (x) Principal Component Analysis (PCA), (xi) Day/Time/Month data decomposition (DTM), (xii) Daylight-free decomposition (DF),

(xiii) Seasonal decomposition (SD), (xiv) Resubstitution validation, (xv) Cross-validation, (xvi) Holdout validation. These DP and model evaluation methods are further classified into six categories as described in Section 4.2. A summary of DP studies can be seen in Table 4-1.

Table 4-1 Study comparisons on data preprocessing methods

DP Methods	Studies						
	Ranjan et al. [9]	Bramm et al. [10]	Phan et al. [11]	Carrera and Kim [12]	Alshdaift et al. [13]	Malvoni et al. [14]	This study
LWDM		✓			✓		✓
LIM					✓		✓
DO		✓	✓				✓
FO	✓						✓
DS	✓						✓
ZS					✓		✓
RM			✓		✓		✓
LLDF	✓		✓	✓			✓
PCC			✓				✓
PCA			✓			✓	✓
DTM			✓				✓
DF		✓		✓			✓
SD							✓
Resub						✓	✓
Cross				✓	✓		✓
Holdout			✓				✓

Note: Cross = Cross-validation, DF = Daylight-free, DO = Delete Outliers, DS = Data Smoothing, DTM = Day/Time/Month, FO = Fill Outliers, Holdout = Holdout Validation, LIM = Linear Interpolation Method, LWDM = List-wise Deletion Method, PCA = Principal Component Analysis, PCC = Pearsons Correlation Coefficient, Resub = Resubstitution validation, RM = Rescaling Method, ZS = Z-Score

According to Table 4-1, no other study identifies and tests these combination of DP models. A further review of the studies shows that there is either a loose or non-existent preprocessing framework that categorizes DP methods based on their use/function for statistical PV forecasting models.

This study will systematically test different combinations of DP techniques on raw meteorological and PV data before they are inputted into an existing statistical PV model. The adjustment by each DP method(s) to the model's forecasting accuracy is recorded and the performance is compared based on RMSE. To achieve this, two scenarios are considered. The first scenario involves comparing the impact of the DP methods mentioned above and their subsequent combination to a base-case scenario.

In this study, the base-case scenario refers to filling or deleting missing data from the original raw data which is sometimes referred to as the most basic level of DP. In the second scenario, the best combination of DP methods from the first scenario is compared to an existing three-step PV forecasting model that uses a combination of RM, LWDM, PCC, and LLDF preprocessing methods. The main contributions of this study are given below:

1. To identify and classify DP methods, relevant to statistical PV forecasting models, into categories that will make scoping additional methods easier. Researchers will be able to easily identify what DP category their method falls under which would help inform an overall strategy for preparing clean data to be used in analysis.
2. To assess the impact of different DP method(s) on raw data by comparing their RMSE performance to base-case scenario.
3. To investigate the impact of combining different DP techniques to improve an existing statistical PV forecasting model, providing an empirical analysis that can be used as a case study for other DP strategies.

The list of DP methods analyzed by this study is not exhaustive. Some DP methods are simply left out due to their irrelevance to the statistical PV model while others are omitted due to research limitations.

This paper is organized as follows. A structured classification of DP methods is presented and described in Section 4.2. Section 4.3 describes the methodology of the three-step forecasting model. Section 4.4 provides parameters for model evaluation including validation methods and error metrics. Section 4.5 presents the case study including data description and analysis flow chart. Results and discussions are described in Section 4.6, followed by the conclusion and recommendations for future research.

4.2 Data preprocessing methods

A DP framework with selected DP methods to be investigated is shown in Figure 4-1. The DP methods mentioned in Section 4.1 are classified into six DP categories

namely: (1) data cleaning, (2) data reduction, (3) data decomposition, (4) data scaling, (5) data integration and (6) data partitioning.

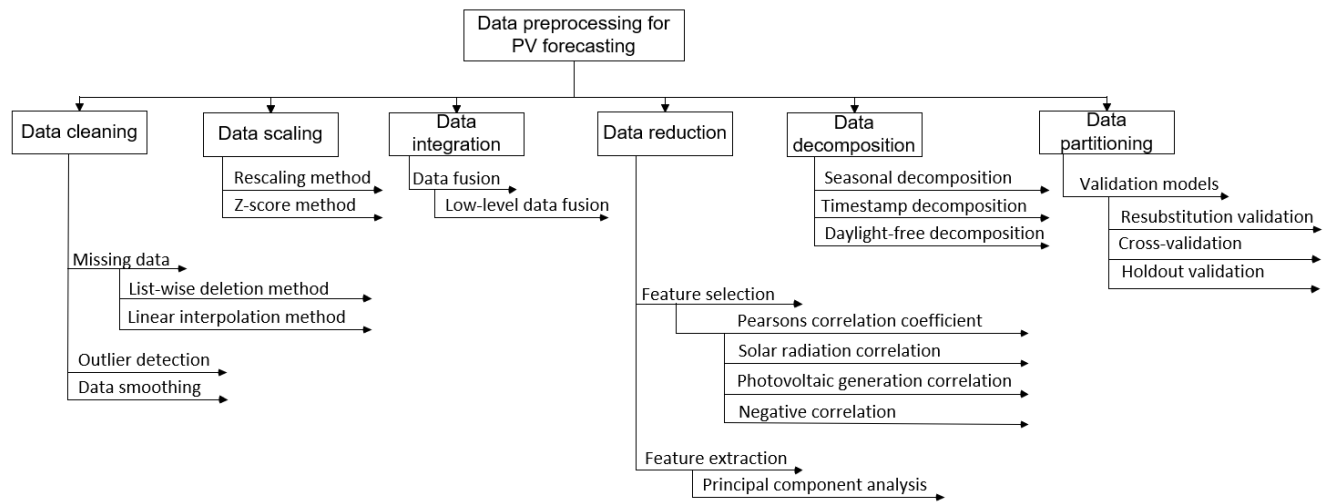


Figure 4-1 Data preprocessing framework for statistical PV forecasting models

4.2.1 Data cleaning

Data cleaning is the process of fixing or eliminating missing data. It involves finding, fixing and/or removing outliers, incorrect, incomplete, or inconsistent data within a dataset [12]. Two DP methods under data cleaning were investigated: handling missing data and outlier detection.

4.2.1.1 Missing data

In compiling large datasets, missing data is commonplace especially if the source of data comes from measuring devices that may malfunction, have communication issues, or are caused by human error. Unaccounted or missing data can lead to a bias in forecasting [13] that can compromise the integrity of any prediction model. Distorted, corrupted or incorrect data may occur during or after data collection and an approach to mitigate their effects on the model development should be considered [14]. Deleting or filling missing data is often the first step in DP and is considered the base-case scenario in this study. In addition, this study used three different sets of data (historical PV generation, weather data from onsite and local weather stations) that have data missing completely at random (MCAR). To address missing data, the Complete Case or Available Case analysis, also known as LWDM, and LIM were used. LWDM is

among the simplest and most common approach to missing data [15]. With this method, if one of the predictor variables has a missing value for a specific time stamp, all measured data points at that time stamp are removed [16]. In other words, LWDM eliminates or excludes a data observation when one or more data variable is missing [15,17]. For instance, in Ziyabari et al. [18], missing and unnecessary data was omitted from a fifteen-year data sample split into ten years for training, three years for testing and two years for validation. Using a multi-branch residual network, Ziyabari et al. [18] were able to achieve a low average RMSE of 0.073 which was better than other networks.

To synthesize the three raw data sources into one, LWDM was used when a row of data equivalent to 1 day was omitted. For example, it was common to have 1 or 2 days MCAR from the PV meter data. In this case, the corresponding timestamps in both the onsite and local weather data sets were deleted. The drawback to this approach is that it leads to a reduced data sample. The advantage is that if the entire data sample is relatively large enough, the missing data does not significantly affect the model, avoiding over approximation while maintaining the integrity of the data sample.

LIM is a heavily used imputation technique [19]. Imputation refers to the process of substituting missing data to create a complete data set [20]. In a study by Kim et al. [21], LIM performed the best in estimating missing environmental and behavioural data. The LIM formula [22] is given in Equation (4-1).

$$y = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)} \quad (4-1)$$

Where x and y are the current points, y_1 , x_1 are the previous points, and y_2 , x_2 are the next points.

4.2.1.2 Outlier detection and data smoothing

Outlier detection identifies data points that differ considerably from the rest of the data or is at odds with an expected data outcome [23]. There are several outlier detection methods like Quartiles, Generalized extreme studentized deviate, Moving averages and so on [24]. Data smoothing refers to methods that eliminate unwanted noise or outliers in the data. The aim is to remove the noise or fluctuations by applying a

statistical or mathematical method while preserving the integrity of the features within the data. The choice of method depends on the data characteristics. A popular outlier detection and data smoothing method is Moving averages or Moving Mean (MM). Moving averages calculate the average of a subset of data points within a sliding or moving window.

4.2.2 Data reduction

Data or dimensionality reduction involves reducing a multidimensional data set into a lower dimension while preserving and retaining the characteristics of the original data set as much as possible [25]. Data reduction comprises of attributes like feature selection, and feature extraction which reduces the dimensionality of the original dataset through methods like PCA.

4.2.2.1 Feature selection

Feature selection process, often referred to as variable selection or attribute selection, involves choosing a subset of relevant and non-redundant features, while maintaining a reduced dimensionality, retaining as much of the original dataset's information as feasible [26]. By reducing the dimensionality of the data set, feature selection can improve the accuracy, effectiveness, speed, generalizability, and interpretability of the forecasting model [27]. The three-step forecasting model uses the PCC analysis to select positive moderate to strong correlated weather parameters to SR. The PCC formula [28] is given in Equation (4-2) and is used to determine which weather parameters have a moderate to strong correlation to PV generated. The greater the correlation coefficient, the better the model [29].

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}} \quad (4-2)$$

Where r is the correlation coefficient, x is one of variables in the sample that is to be compared, y represents the other variable in the sample, $\bar{x}$ and $\bar{y}$ are the mean values of the respective variables.

Like the proposed model, several studies [30–34] use PCC in selecting features. In Zeng et al. [35], PV generated is correlated with more than ten parameters while in Mansood et al. [33], PV is correlated with five parameters. In this study, solar radiation

correlation (SRC) and PV correlation (PVC) are both investigated to determine which parameter is better at selecting relevant features.

In addition, the selection of negative correlated features is also investigated. Negative correlation (NC) describes an inverse relationship between two variables where one variable increases as the other decreases [36]. NC has important applications in many prediction models. For instance, heating energy demands increase as temperature decreases. In a review of solar irradiation forecasting methods, Patel et al. [37] describes an inverse relationship with humidity and classifies it as one of the weather parameters used for model training. This inverse relationship is synonymous with NC. In Kong et al. [34], the R-value benchmark for input selection was set at 0.25 for seven weather parameters. Humidity, with an R-value of -0.5681, was selected as one of the inputs.

4.2.2.2 Feature extraction

PCA is a common technique used in feature extraction and dimension reduction [38]. It reduces the number of variables in a system by expressing the variables as linear combination of each other. It does so by identifying the driving principles in a variable that may exist in other variables in the same system. It then factors out this common principle and removes any redundancy or repetition. This way, the variable size becomes smaller, and it is easier for the forecasting algorithm to detect patterns and identify relationships in the data. The new set of variables are referred to as the principal components where each component is a linear combination of the original variables. It is important to note that all principal components are orthogonal to each other meaning there is no redundant information. It is also important to note that no data characteristic is eliminated, and the full set of principal components is as large as the original set of variables. Several studies [39–44] use PCA as a preprocessing technique to reduce the dimensionality of input data to enhance forecasting accuracy and model performance. In Kazem et. al [39], PCA produced better PV power forecasting results having a mean square error (MSE) of 0.0780 versus Full Recurrent Neural Network (FRNN) which had 0.1778. PCA is also used as a feature selection technique in a study by Munawar et al. [44], as feature selection has a significant impact on model performance. The study used the XGBoost model with PCA and this produced the lowest RMSE. In addition, PCA outperformed the feature importance

method for feature selection. Mallika et al. [43], introduces a novel ionospheric forecasting method to forecast the ionospheric total electron content by combining PCA and artificial neural networks (ANN). This PCA combined model outperformed normal neural networks with a correlation of 0.95 as opposed to 0.41 for the latter. In another study by Lan et al. [40], the research incorporates Discrete Fourier Transform (DFT), PCA and Elman Neural Network (ENN) to track changes in the frequency of SR to provide a more accurate SR forecast. To investigate accuracy performance of various DP techniques on a time series forecasting model by Malvoni et al. [41], the approximation and detail coefficients resulting from analysing meteorological data inputs using PCA and Wavelet Decomposition, resulted in a decrease in prediction error of up to 1% over other methods. To determine daily SR forecasts, Zhang and Wei [42] used PCA to reduce dimensions of decomposed historical SR. To test the effectiveness of PCA, a hybrid model based on Wavelet Transformation, Extreme Learning Machine and Bat Algorithm was used with and without PCA. A conclusion of the study was that the model performed far better when PCA was enabled.

PCA in MATLAB uses the singular value decomposition algorithm. Let X be a real $m \times n$ matrix of rank n , and I_n is the identity matrix. A singular value decomposition of X [45] is such that:

$$X = U\Sigma Z^T \quad (4-3)$$

Where: U is a $m \times r$ matrix called the left singular values of X and $U \times U^T = I_r$

Σ is an $r \times r$ matrix that is diagonal with single values of X

Z is a $n \times r$ matrix called the right singular value of X and $Z \times Z^T = I_r$

Using singular decomposition,

$$T = XZ \quad (4-4)$$

$$T = U\Sigma Z^T Z \quad (4-5)$$

$$T = U\Sigma \quad (4-6)$$

The output of PCA is the pair of matrices U and Σ . Matrix T is now comparable to the original matrix X such that each column of T is a left singular vector of X multiplied by the corresponding singular value.

4.2.3 Data decomposition

Data decomposition is used to break down the original time series data into a number of comparatively steady subseries [46]. The aim is to identify and group each subseries by their unique characteristics, making it easier for the prediction algorithm to identify patterns in the data. For example, data decomposition can identify seasonality and trends in a time series data. A forecasting model is built on each subseries to obtain several individual forecasting outcomes. The final forecasting result is obtained by totalling up all the individual forecasting outcomes. Examples of data decomposition models are Wavelet Decomposition (WD), Atomic sparse decomposition (ASD), Empirical mode decomposition (EMD), and Variational mode decomposition (VMD) [47].

Classifications of decomposition include additive and multiplicative decomposition. Seasonal decomposition, an example of additive decomposition, breaks down a time series into its various components, including trend, seasonality, and remainder (residuals or noise) while capturing seasonal characteristics in the observed data [48]. Data can also be grouped by days of the month, time and even the months in a year in a timestamp decomposition. The logic behind dividing data into seasons is that days in a particular season tend to have similar weather patterns. By grouping similar weather days together and training the model, the machine learning algorithm should be able to detect patterns easily. Several authors have explored this concept. Fraihat et al. [49] investigates the influence of twenty-four (24) weather parameters on SR forecasting in the western region of Jordan. The study concludes that the parameters impact differently on SR depending on the season. In Singla et al. [50], a hybrid model is investigated which is a combination of Full Wavelet Packet Decomposition (FWPD) and Bidirectional Long Short-Term Memory (BiLSTM). In this study, the seasons are divided into, Summer (March-May), Monsoon (June – September), Autumn (October-November), and Winter (December-February) according to the Indian Meteorological Department. Results showed that there were lower forecasting errors in the summer and winter than in the Autumn and Monsoon. Ali-Ou-Salah et al. [51] presented another hybrid approach based on ANN and seasonal clustering for ultra-short term global solar radiation forecasting. In yet another study on solar power forecasting, the hybrid approach, consisting of Empirical Wavelet Transform (EWT) and Robust Minimum Variance Random Vector Functional

Link Network (RRVFLN) with random weight vector was investigated. The study acknowledges the fact that factors like seasons and geographical area affects solar power. The data site is located in Alabama and months were split into Spring (March-May), Summer (June-August), Autumn (September-November) and Winter (December-February).

Data can also be classified based on trends like the hours of PV generation. PV generation mirrors the solar intensity for a specific location and follows a bell curve pattern when plotted against the 24 hours in a day. This means that certain hours in the morning and at night will have zero PV generation because the sun is no longer available. These hours are referred to as DF hours. Some publications have used the method of excluding daylight-free hours in their analysis [50,52–56].

4.2.4 Data scaling

Data scaling, data transformation or data normalization is a process that entails changing the scale of the variables or features in your dataset to a similar scale. The purpose is to prevent certain features with larger values from dominating others and to ensure they all contribute equally to the model [57]. Common methods of data scaling include Min-Max scaling, Z-score normalization, Robust scaling, Log transformation.

The Min-Max method is also known as the RM. With a typical range of [0, 1], RM prevents over-influencing of one over the other due to differing lengths. The RM equation [58] is given in Equation (4-7).

$$X_{rescaled} = a + \left[\frac{X - X_{min}}{X_{max} - X_{min}} \right] (b - a) \quad (4-7)$$

Where $X_{rescaled}$ is the normalized data between the intervals [a, b], X represents the values of the x-variable in the sample, X_{min} is the minimum while X_{max} is the maximum value of X in the sample.

ZS or the Standard Score normalization method is another way of normalizing data. Several studies [49,52,59–62] highlight the use of ZS as a preferred normalization method. The ZS equation [63] is defined in Equation (4-8) where μ and σ are the mean and standard deviations of the sample respectively.

$$Z = \frac{X - \mu}{\sigma} \quad (4-8)$$

4.2.5 Data integration

Data integration is the process of combining data from different sources [64] to enhance the users view of the data. The data combined must be accurate, consistent, and of a similar format. Data fusion falls under the broader umbrella of data integration. Data fusion methods combine information gathered from multiple sources to extract meaningful information [65]. Data fusion makes it possible to obtain more informative and complete representation of the data than would be possible by just using one data source. In general, data fusion can be classified as low, intermediate, or high level based on the processing stage at which information fusion occurs [66].

LLDF is a simple technique that combines or concatenates raw data from more than one source into a single data sample ready for data analysis [67]. In a study by Teglia et al. [68], a LLDF strategy was proposed for data processing of second-order liquid chromatographic data. The LLDF strategy was successful in overcoming baseline contribution, chromatographic peaks distortions and peak overlaps while increasing the global prediction ability and decreasing uncertainty. In another study, Lanza et al. [69] prove that the LLDF approach provided greater discriminating capacity of milk than either of the individual fatty acid-nonpolar and the nuclear magnetic resonance polar data sets, providing a better predictive performance of the Canonical discriminant analysis algorithm.

4.2.6 Data partitioning

Data partitioning or data splitting is the process of dividing a dataset into subsets or partitions for a variety of uses, including training machine learning models, validating model performance, and testing model generalization. Traditional data splitting involves dividing the data into training and testing data sets for training and evaluating the model respectively [70,71]. In this study, 75% of data is used for training while the remaining 25% data is used for testing.

4.3 Three-step forecasting model

The existing three-step statistical PV forecasting model [72] uses positive correlated weather parameters to solar radiation (SR), from an on-site weather station and a local weather station, to predict PV generation. The choice of SR was determined because it has the highest correlation to PV generation of any other parameter [73]. The raw data was preprocessed using LLDF, PCC, LWDM and RM. Resubstitution was arbitrarily selected as the validation method. Figure 4-2 shows the existing forecasting architecture.

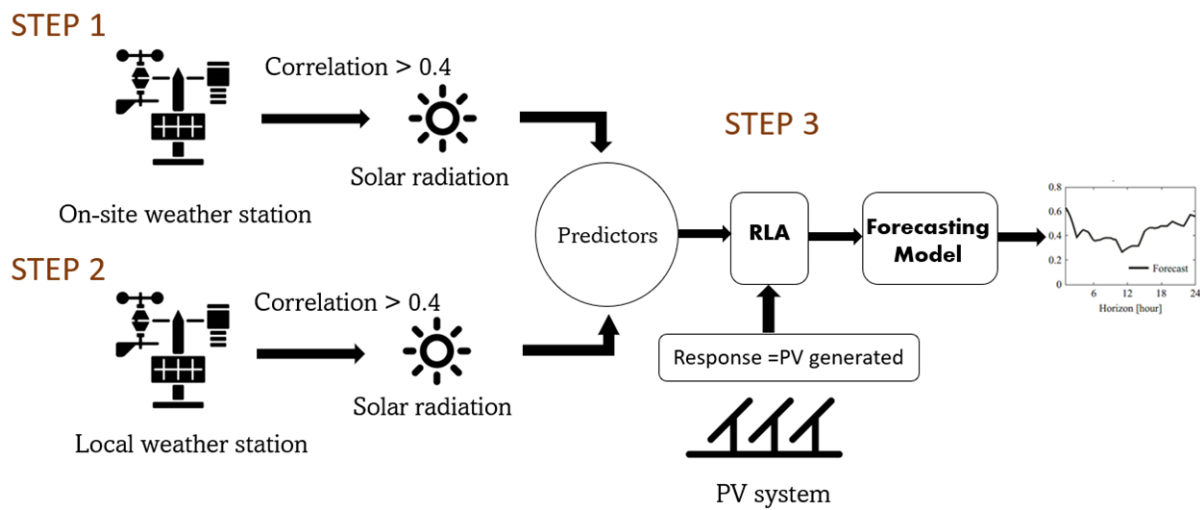


Figure 4-2 Three-Step Forecasting Architecture

4.4 Model evaluation

This involves the tools and techniques used in assessing the performance of the model.

4.4.1 Validation methods

Validation methods are essential when developing the model and define what percentages of the data sample is used for training, testing, and validating a model. These methods prevent overfitting or underfitting of the model by testing the forecasting algorithm on new data. It achieves this by also partitioning the data into training and testing data subsets. MATLAB has three validation method options: Cross-validation, Holdout validation and Resubstitution validation.

In Cross-validation, the data set is divided into folds or division of roughly equal size. A fold or subset is used to validate or test the model, while the remaining subsets are used to train the model. This process is repeated until each subset has been used exactly once for validation. The average of the validation error over all the folds is determined and this becomes the error for the prediction model. Holdout validation separates the data into training and validation sets based on a percentage. The advantage to this is that the data is trained and validated only once and doesn't have to be repeated unlike in cross-validation. There is however a caution when this method is used for small samples. Resubstitution validation doesn't partition the data but instead uses all the data to compute the error. Without any separate validation data, there is no protection against overfitting using this method.

In Zhang et al's study [74], both resubstitution and cross-validation methods were used. Worthy of note in the study was the use of cross-validation to reduce and correct resubstitution bias that arose when designing confirmatory clinical trials using a finite sample. In a study for predicting the occurrence of respiratory diseases using climatic and air-pollution factors by Ku et al. [75], holdout validation was performed using 30% of the data when training gradient boosting and Gaussian Process Regression (GPR) machine learning models. Similarly, cross-validation has been used in several studies [39,44]. Mallika et al. [43] used cross-validation with $k = 5$ for feature selection and $k = 3$ for hyperparameter selection in forecasting the ionospheric space weather. In a comparative study of classification tools [76], all three validation methods were used. The data was trained using five holdout percentages (60%, 65%, 70%, 75%, 80%) and $k = 10$ folds for cross-validation.

4.4.2 Performance metrics

There are quite a few metrics to measure the performance of a forecasting model. These error equations show the difference between the actual data and the forecasted or predicted data. The following error formulas [77,78] are used to evaluate the performance of the model:

- I. **MSE:** Is the average of the squares of the errors or the sum of the average difference squared between predicted and actual values.

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2 = RMSE^2 \quad (4-9)$$

- II. **MAE:** Measures the average absolute difference of the forecasted values from the actual values.

$$MAE = \frac{1}{N} \sum_{i=1}^N |(\hat{x}_i - x_i)| \quad (4-10)$$

- III. **RMSE:** Penalizes big forecast errors and is preferably used where small errors are acceptable [54]. It is calculated by taking the square root of the MSE. The lower the RMSE, the higher the prediction model accuracy.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2} \quad (4-11)$$

- IV. **R-Squared:** The coefficient of determination or R-squared indicates the proportionate amount of variation in the dependent variable response explained by the independent variable. In other words, R-squared is a goodness of fit measure that shows how well the data fits the regression model [37].

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{x}_i - x_i)^2}{\sum_{i=1}^N (\hat{x}_i - \hat{x}_{average})^2} \quad (4-12)$$

Where N is the number of data points, $\hat{x}_i$ are the predicted values and x_i are the observed values.

Using a combination of MSE, MAE, R-squared, and RMSE offers a comprehensive evaluation of forecasting models. While MSE is valuable for its sensitivity to large errors (by squaring them), MAE provides robustness to outliers and errors by giving them equal weights, and clear interpretability by being expressed in the same units as

the original data. RMSE retains the interpretability of MAE and sensitivity to error of MSE, making it a versatile and widely accepted metric. R-squared is useful for comparing different models on the same dataset, as a higher R-squared indicates a better fit. Together, these metrics ensure a thorough assessment of model performance, considering both the magnitude of errors and their distribution, which is crucial for robust and reliable forecasting in practical applications.

4.5 Case study

Regression analysis is used to determine the relationship and future predictions of a dependent variable (the PV generated) from a number of independent variables (weather parameters). The Regression Learner App (RLA) in MATLAB, through the lowest RMSE, selects the most accurate out of more than 20 regression models. Before training via the regression models, the response (independent variable) and predictors (dependent variables) must be selected. In this study, two and half years (2018 – 2020) of historical data comprising of nine weather parameters from the Kisanhub weather station (KH075) [79] located in Cranfield University and twelve weather parameters from The Metrological Office in Bedford, U.K are provided. The corresponding historical PV data is obtained from Cranfield's online energy manager platform [80]. To investigate the impact of the DP methods highlighted in Section 4.2, the RLA flowchart is used and shown in Figure 4-3. In addition, the methods are systematically investigated in the order they are presented in Figure 4-1 from left to right with the results from each trial displayed in Table 2. The default validation method used is resubstitution. The aim is to determine which combination of DP methods significantly improves the base-scenario and the accuracy of the forecasting model.

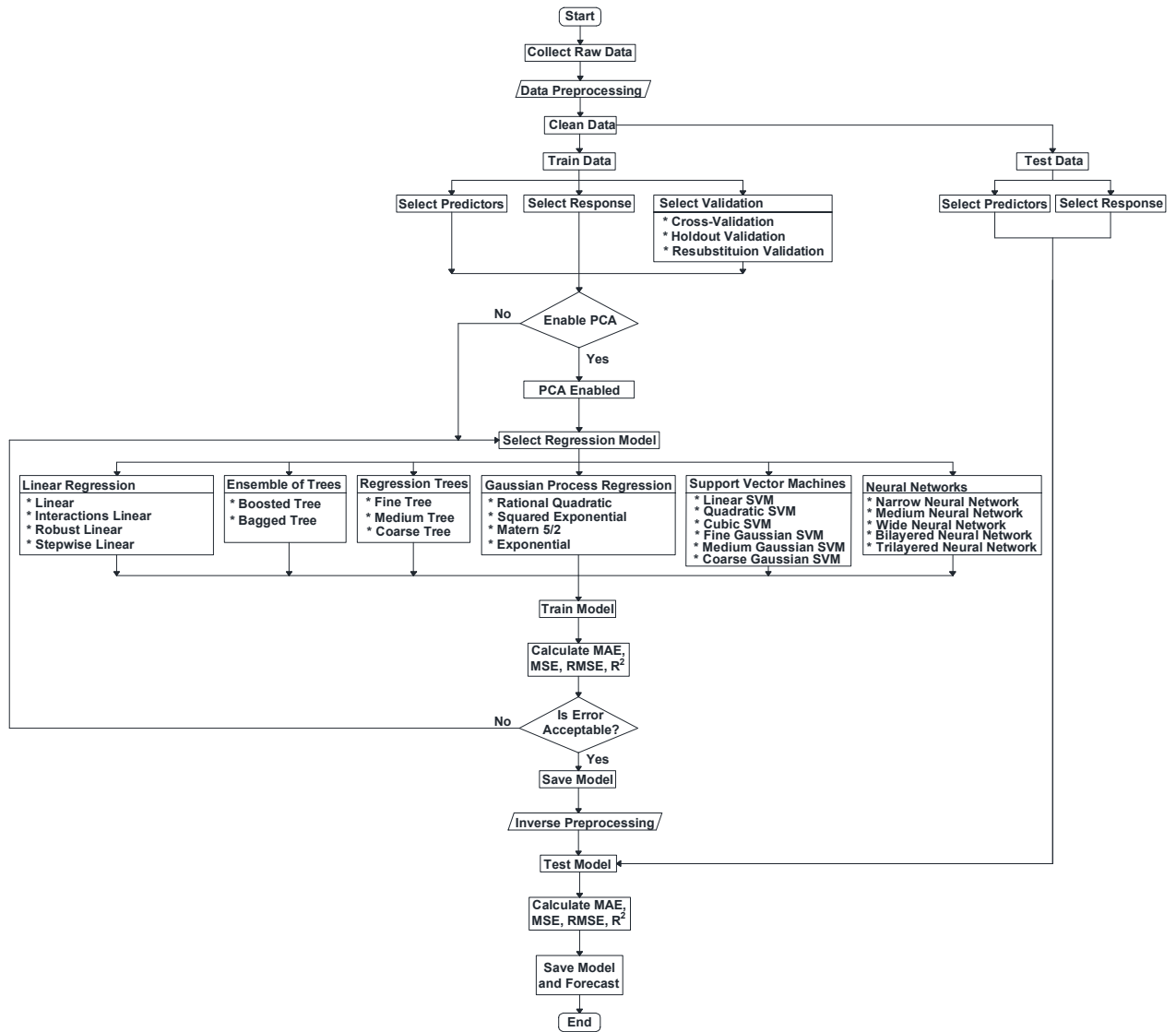


Figure 4-3 Modified flow chart of the Regression learner app

4.6 Results and discussion

Results from the test data based on the lowest RMSE values and the best performing algorithm for each trial are displayed in Table 4-2. The lowest RMSE for each sub-group of DP methods have been highlighted.

Table 4-2 Performance of data preprocessing methods

Data cleaning													
Trial	LWDM	LIM	DO	FO	DS	RMSE	Algorithm						
1	✓					38.52	MGSVM						
2	✓		✓			35.79	CGSVM						
3	✓			✓		38.40	MGSVM						
4	✓				✓	11.39	WNN						
5	✓		✓		✓	12.03	WNN						
6		✓				37.56	MGSVM						
7		✓	✓			37.15	MGSVM						
8		✓		✓		38.09	MGSVM						
9		✓			✓	11.53	WNN						
10		✓	✓		✓	12.74	WNN						
Data scaling													
	LWDM	LIM	DS	ZS	RM	RMSE	Algorithm						
11	✓		✓	✓		0.2044	MNN						
12	✓		✓		✓	0.0403	WNN						
13		✓	✓	✓		0.1745	WNN						
14		✓	✓		✓	0.0375	WNN						
Data integration													
	LIM	DO	FO	DS	RM	Cranf	Bed	Cranf + Bed	RMSE	Algorithm			
15	✓			✓	✓	✓			0.0422	FGSVM			
16	✓			✓	✓		✓		0.0530	FGSVM			
17	✓			✓	✓			✓	0.0375	WNN			
18	✓	✓		✓	✓			✓	0.0379	WNN			
19	✓		✓	✓	✓			✓	0.0295	MGSVM			
Data/Dimentionality reduction – Feature selection by correlation (SRC,PVC, NC)													
	LWDM	LIM	FO	DS	RM	Cranf	Bed	Cranf + Bed	SRC	PVC	NC	RMSE	Algorithm
20	✓				✓	✓			✓			0.1037	BNN
21	✓		✓		✓	✓			✓			0.1042	NNN
22	✓		✓	✓	✓	✓			✓			0.0604	WNN

23	✓				✓		✓		✓			0.1049	TNN
24	✓		✓		✓		✓		✓			0.1073	NNN
25	✓		✓	✓	✓		✓		✓			0.0688	FGSVM
26	✓				✓			✓	✓			0.0945	MNN
27	✓		✓		✓			✓	✓			0.0928	BNN
28	✓		✓	✓	✓			✓	✓			0.0549	FGSVM
29		✓	✓	✓	✓			✓	✓			0.0547	WNN
30		✓	✓	✓	✓			✓		✓		0.0417	WNN
31		✓	✓	✓	✓			✓	✓		✓	0.0362	FGSVM
32		✓	✓	✓	✓			✓		✓	✓	0.0365	FGSVM
Data decomposition – Day, Time , Month, Daylight-free													
	LIM	FO	DS	RM	Cranf + Bed	SRC	PVC	NC	DTM	DF		RMSE	Algorithm
33	✓	✓	✓	✓	✓	✓		✓	✓			0.0264	WNN
34	✓	✓	✓	✓	✓		✓	✓	✓			0.0251	WNN
35	✓	✓	✓	✓	✓		✓	✓	✓	✓		0.1984	WNN
Data decomposition – Seasons													
	LIM	FO	DS	RM	Cranf + Bed	PVC	NC	DTM	Season			RMSE	Algorithm
36	✓	✓	✓	✓	✓	✓	✓	✓	Summer			0.0128	RQGPR
37	✓	✓	✓	✓	✓	✓	✓	✓	Spring			0.0144	MGPR
38	✓	✓	✓	✓	✓	✓	✓	✓	Autumn			0.0090	SEGPR
39	✓	✓	✓	✓	✓	✓	✓	✓	Winter			0.0079	MGPR
Data/Dimentionality reduction – PCA													
	LIM	FO	DS	RM	Cranf + Bed	PVC	NC	DTM	PCA			RMSE	Algorithm
40	✓	✓	✓	✓	✓	✓	✓	✓	PCA 11			0.0265	WNN
41	✓	✓	✓	✓	✓	✓	✓	✓	PCA 9			0.0264	WNN
42	✓	✓	✓	✓	✓	✓	✓	✓	PCA 8			0.0253	WNN
43	✓	✓	✓	✓	✓	✓	✓	✓	PCA 7			0.0276	WNN
44	✓	✓	✓	✓	✓	✓	✓	✓	PCA 5			0.0522	FGSVM
Data paritioning													
	Validation		PCA		RMSE	Algorithm		PCA	RMSE	Algorithm			
45	Cross : 5		9		0.0280	MG SVM		8	0.0253	WNN			
46	Cross: 10		9		0.0272	WNN		8	0.0254	WNN			

47	Cross: 15	9	0.0265	WNN	8	0.0248	WNN
48	Cross: 20	9	0.0268	WNN	8	0.0268	WNN
49	Holdout: 10	9	0.0250	WNN	8	0.0253	WNN
50	Holdout: 15	9	0.0250	WNN	8	0.0248	WNN
51	Holdout: 20	9	0.0250	WNN	8	0.0256	WNN
52	Holdout: 25	9	0.0253	WNN	8	0.0266	WNN

Data decomposition – Seasons

	Seasons	PCA	Validation	RMSE	Algorithm	Validation	RMSE	Algorithm
53	Spring	8	Cross: 15	0.0142	MGPR	Holdout: 15	0.0115	MGPR
54	Summer	8	Cross: 15	0.0117	RQGPR	Holdout: 15	0.0116	RQGPR
55	Autumn	8	Cross: 15	0.0078	MGPR	Holdout: 15	0.0077	MGPR
56	Winter	8	Cross: 15	0.0076	MGPR	Holdout: 15	0.0077	MGPR

Note: Bed = Bedford, BNN = Bilayered Neural Network, Cranf. = Cranfield, Cross : # = Cross-validation : Number of folds, CSVM = Coarse Gaussian Support Vector Machine, DF = Daylight-free, DO = Delete Outliers, DS = Data Smoothing, DTM = Day-Time-Month, FGSVM = Fine Gaussian Support Vector Machine, FO = Fill Outliers, Holdout = Holdout Validation : Percent hold out, LIM = Linear Interpolation Method, LWDM = List-wise Deletion Method, MGPR = Matern 5/2 Gaussian Process Regression, MGSVM = Medium Gaussian Support Vector Machine, MNN = Medium Neural Network, NC = Negative Correlation, NNN = Narrow Neural Network, PCA = Principal Component Analysis, PVC = Photovoltaic Correlation, RM = Rescaling Method, RMSE = Root Mean Square Error, RQGPR = Rational Quadratic Gaussian Process Regression, SEGPR = Squared Exponential Gaussian Process Regression, SRC = Solar Radiation Correlation, TNN = Trilayered Neural Network, WNN = Wide Neural Network, ZS = Z-Score

The data from the on-site weather station, the PV meter, and the local weather station are recorded in 15-minute, 30-minute and 1-hour timestamps respectively. The first step was to synchronize the data sampling rates, ensuring that only data values with the same corresponding hourly timestamps are used. The PV meter had several days missing at a time. These along with the corresponding observations from the weather station were removed to have an equal number across all observations and to avoid over approximation. Using LLDF, the raw data from the two weather stations was combined to make cleaning the data easier. The result was a raw data set of 21 weather variables (excluding time and date) with hourly data points from 18/07/2018 – 31/12/2020 numbering 436,527 individual data points and 20,787 hourly observations.

In investigating the DP methods, the first step is to clean the data. To handle missing data, two methods are considered: LWDM and LIM. In LWDM, incomplete or missing data was deleted. In other words, if any variable had a missing hourly data point, the entire row of data containing the hourly readings for the other 20 variables was deleted. 3053 observations were deleted representing a decrease of 15% from the original raw data set. In LIM, missing hourly data is filled by linear interpolation using Equation (4-1). In Table 4-2, Trial 1 and Trial 6 show LWDM with an RMSE of 38.52 and LIM with an RMSE of 37.56 respectively, which represent some of the highest RMSE values of all the trials. LWDM and LIM are often considered the base-case scenario as suggested in Section 4.2.1.1 and are a reference by which all the other DP methods are measured. Trial 1 to 10 show the test results of LWDM and LIM in combination with deleting outliers (DO), filling outliers (FO) and/or data smoothing (DS). MM, with a centred moving window of length 11 and threshold factor of 3, was used in both FO and DS. Of all the trials, the combination of either LWDM + DS or LIM + DS had the lowest RMSE results with 11.39 and 11.53 respectively. RMSE of the test results were calculated using Equation (4-9). Furthermore, DS significantly improved the accuracy in all the tests it was featured in. This is further validated in trials 20-22 where the addition of DS in trial 22 produces the lowest RMSE (0.0604) of the three trials. DS is therefore a significant tool used in improving model accuracy. Wide Neural Network (WNN) and Medium Gaussian SVM (MG SVM) were consistently the best performing algorithms probably. This may be because Support Vector Machines (SVM) use of kernels allows SVMs to implicitly handle non-linear

relationships between features without explicitly transforming the data [81], while Neural Networks (NN) use of weight adjustment allows neural networks to adapt to the scale of input features during training [82]. Additional DP methods need to be investigated to establish a clear-cut pattern.

In data scaling, two normalization techniques are investigated: RM and ZS. Before training the data, normalization was performed using Equations (4-7) and (4-8). In general, all the trials from 11 - 56 performed better and had significantly lower RMSE values than tests from 1 - 10 because the data had been normalized. By including DS based on previous results, the combinations of LWDM + DS and LIM + DS are investigated separately after normalization using either RM or ZS. RM outperformed ZS in all the tests. Although close, the combination of LIM + DS + RM achieved better results (0.0375) than LWDM + DS + RM (0.0403), further validating LIM over LWDM.

Data integration involves testing LLDF by combining raw data from multiple sources. Three model training scenarios are investigated based on the data source; data from Cranfield alone, data from Bedford alone, and data from both weather stations combined using LLDF. Trial 17 using LLDF to combine data from Cranfield and Bedford outperformed the other data sources. Importantly, the impact of DS and RM was confirmed and supports the conclusion reached in the previous trials. Furthermore, the addition of FO outperformed DO resulting in RMSE values of 0.0295 and 0.0379 respectively.

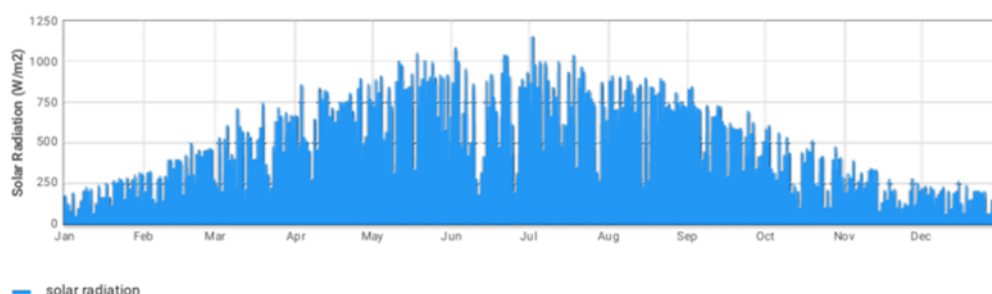
The next series of tests explores data reduction in terms of correlation. In the original three-step forecasting architecture, feature selection was based on weather parameters with moderate to strong positive correlation (>0.4) to SR. These next trials will investigate the impact of replacing SR with PV generated while including weather parameters that have a strong NC. To select these weather parameters, PCC analysis on the data cleaned by LIM is performed using Equation (4-2). The results are displayed in Table 4-3.

Table 4-3 Correlation matrix of weather Parameters in Cranfield and Bedford weather stations to ((a) Solar radiation (b) Photovoltaic generation)

<i>SRC</i>		<i>PVC</i>	
Solar Radiation	1	PV	1

Wind Speed	0.0175	Solar Radiation	0.9444
Relative Humidity	-0.5961	Wind Speed	0.0251
Pressure	0.1973	Relative Humidity	-0.5026
Dew Point	0.2025	Pressure	0.2144
Heat Index	0.5247	Dew Point	0.1511
Temperature	0.5464	Heat Index	0.4243
Wind Chill	0.5412	Temperature	0.4427
Hourly Pressure at Mean Sea Level	0.1902	Wind Chill	0.4383
Hourly Dewpoint Temperature	0.2014	Hourly Pressure at Mean Sea Level	0.2085
Hourly Rainfall Total	-0.1544	Hourly Dewpoint Temperature	0.1331
Hourly Relative Humidity	-0.6979	Hourly Rainfall Total	-0.1634
Hourly Mean Wind Direction	-0.0611	Hourly Relative Humidity	-0.6249
Hourly Mean Windspeed	-0.0127	Hourly Mean Wind Direction	-0.0458
Hourly Maximum Gust	0.0287	Hourly Mean Windspeed	-0.1091
Hourly Temperature	0.5437	Hourly Maximum Gust	0.0214
Hourly Wet Bulb Temperature	0.4092	Hourly Temperature	0.4443
Hourly Global Radiation	0.9504	Hourly Wet Bulb Temperature	0.3215
Hourly Total Cloud Cover	-0.2850	Hourly Global Radiation	0.8962
Hourly Visibility	0.2397	Hourly Total Cloud Cover	-0.3406
		Hourly Visibility	0.2017

The highlighted variables in Table 4-3 shows that the number of features has been reduced from 20 to 9 using SRC, and from 21 to 8 using PVC, with the hourly wet bulb temperature being the only difference. This is because the correlation value is now less than 0.4 and cannot be considered as a predictor when PV is used as the determining parameter. In a comparison of SRC and PVC, PVC outperforms SRC with an RMSE value of 0.0417 to 0.0547 respectively. According to the correlation matrixes in Table 4-3, Relative Humidity is the only weather parameter that has a moderate to strong negative correlation to both SR and PV. Figure 4-4 clearly shows an example of this inverse relationship using 2019 weather data from the Kisanhub weather station.



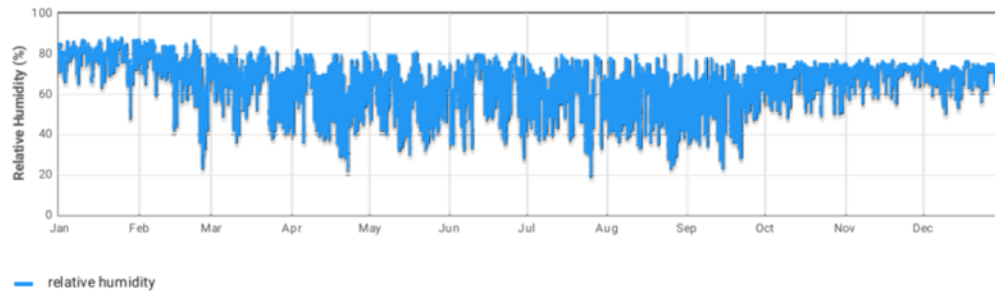


Figure 4-4 Weather data from Kisanhub, Cranfield University showing the inverse relationship between Solar Radiation and Relative Humidity in 2019

By adding weather parameters with moderate to strong NC, both SRC and PVC results mentioned above are further improved resulting in RSME results of 0.0362 and 0.0365 respectively. Table 4-4 shows a summary of the training and testing results with errors and model fit calculated by using Equations (4-9 - 4-12).

Table 4-4 Performance of Solar Radiation correlation and Photovoltaic correlation

SRC	Training				Testing			
Model	RMSE	MSE	MAE	R²	RMSE	MSE	MAE	R²
FGSVM	0.0275	0.0008	0.0172	0.97	0.0362	0.0013	0.0571	0.95
PVC	Training				Testing			
Model	RMSE	MSE	MAE	R²	RMSE	MSE	MAE	R²
FGSVM	0.0279	0.0008	0.0173	0.97	0.0365	0.0013	0.0237	0.95

According to the results of Table 4-4, SRC and PVC have relatively close RMSE test results. As further evidence of this, Figure 4-5 shows the scatter plots of the two methods with similar variations along the regression line.

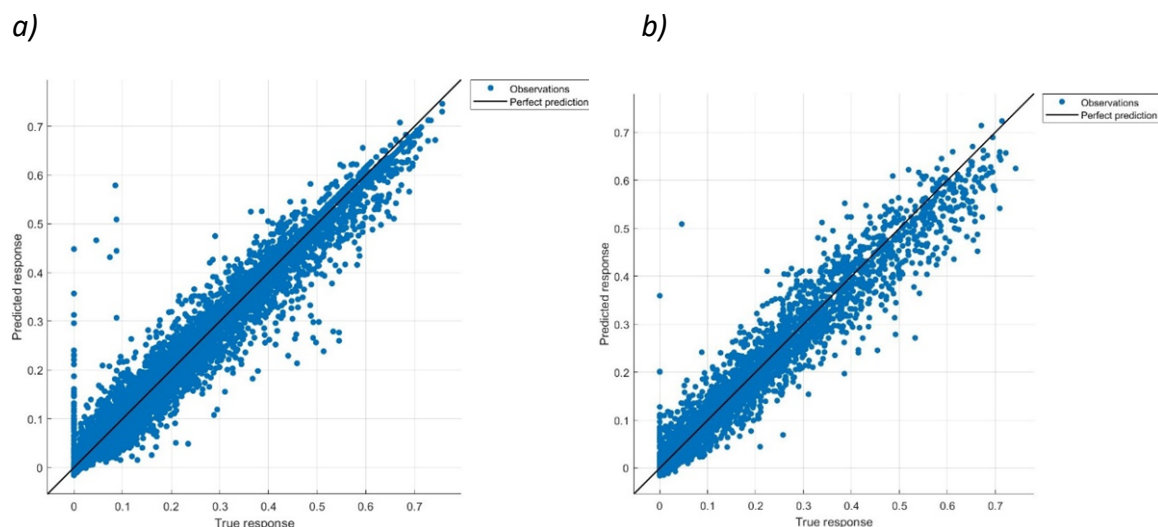


Figure 4-5 Model Performance of ((a) Solar Radiation correlation (b) Photovoltaic correlation)

This concludes that PV can be used instead of SR in the forecasting model. The result of which is the better performing method is inconclusive and more tests need to be carried out. In reviewing trials 20 – 32, results of DP method performances from previous trials were further validated. In trials 22, 25, and 28, the addition of DS produced better results than the previous trials without DS. In addition, models that used LLDF in the test trials (26 -32), performed better than models using only one source of data (20 -25). The LIM models in trials 29 – 32, outperformed all the trials using LWDM (20 – 28).

To investigate data decomposition, elements of date are added including dividing the data set into seasons and considering DF hours. Timestamp decomposition involves breaking the data timestamp into day of the month, time of the day, and month of the year (DTM), adding them as three unique features to the SRC and PVC predictors. In trials 33 -34, PVC with DTM outperforms SRC with DTM resulting in RMSE values of 0.0251 and 0.0264 respectively. With this result, it is concluded that the PVC method is slightly better than SRC. DF hours are investigated by removing the hours of zero PV production from the data set. The 0.1984 RMSE value in trial 35 shows that DF reduces the prediction accuracy. A possible reason behind this is by eliminating DF hours, other important features that help in training and testing of the data are eliminated as well. The more useful information is deleted from the data set, the less accurate the prediction is. To investigate how the model performs in seasons, the data was divided into Spring (March 20th to June 20th), Summer (June 21st to September 21st), Autumn (September 22nd to December 20th) and Winter (December 21st to March 19th) [83]. The RMSE values in trial 36 – 38 are the lowest of all the trials. An explanation is that the training and testing of the models are performed over relatively short periods of time. The relative short time and the fact that the data is largely homogenous leads to relatively low RMSE values. A further observation from these test trials is that the best performing algorithms in training and testing of the seasonal data came from the GPR family of algorithms. GPR are non-parametric, non-linear probabilistic kernel-based regression functions [84]. With the adaptation of the

kernel-based function, GPR show wider adaptability in dealing with multi-dimensional and complex nonlinear forecasting problems [85]. Based on the Bayesian framework [86], GPR computes the probability of an input vector that follows the Gaussian distribution and is expressed in the form of its mean and covariance [87].

Impact of PCA is also considered. The component reduction criteria of the PCA analysis in MATLAB allows the user to select how many components or combination of components should be used. Currently, the total number of features for the PVC-based model which was originally 8 is now 11 after adding DTM. Using Equations ((4-3) - (4-6)), trials 40 - 44 test the addition of PCA starting from 11 components all the way down to 5 components. Trial 42 which represents PCA with 8 components (PCA 8), has the lowest RMSE of all the other components at 0.0253 with PCA 9 coming in at a close second at 0.0264. A possible explanation could be that of all the 11 features, relative humidity, temperature, and SR occur twice. By combining these doubling features together, we are left with 8 components. More importantly is that there doesn't seem to be an added advantage over the previous trial (34) without PCA. This might be due to a relatively small sample size or non-linearity in the data [88]. However, additional tests will be conducted.

The default validation method for all the trials has been resubstitution. In trials 45 - 52, different folds and percentages were investigated for cross and holdout validation methods respectively while still maintaining all the DP methods from trials 40 - 44. The choice of folds started at 5 and stopped at 20 when the RMSE results began to increase. For holdout validation, the partition percentages stop at 25% because after then (at 30%), the test sample size drops from 25% to 20%, which is inconsistent with the parameter set for this study. Overall, similar RSME values are obtained, with cross validation using 15 folds and holdout validation with 15% having the best results at 0.0248. The similar performance over the three validation methods indicates that the data sample size is adequate and within range of their capabilities. However, resubstitution creates bias and overfitting in smaller data samples. It is recommended to use cross validation for smaller data samples and holdout validation for larger data sets. On closer inspection, PCA 8 performs better in 4 out of the 8 trials and ties with PCA 9 in 1 of the trials. Holdout validation of PCA 9 results has slightly performed better on average over the four percentage partitions (10% to 25%). The seasonal decomposition is repeated, comparing cross validation and holdout validation. In trials

53-56, holdout validation performs better and produces the lowest RMSE figures of all the 56 trials. In addition, the GRP models were the best performing in seasonal decomposition supporting the findings earlier in trials 36 - 39. In summary, the optimal combination of DP methods (LLDF + LIM + FO + DS + RM + PVC + NC + DTM+ PCA 8 + Holdout: 15) resulting in an RMSE of 0.0248 is shown in Figure 4-6 and highlighted in bold as the critical path.

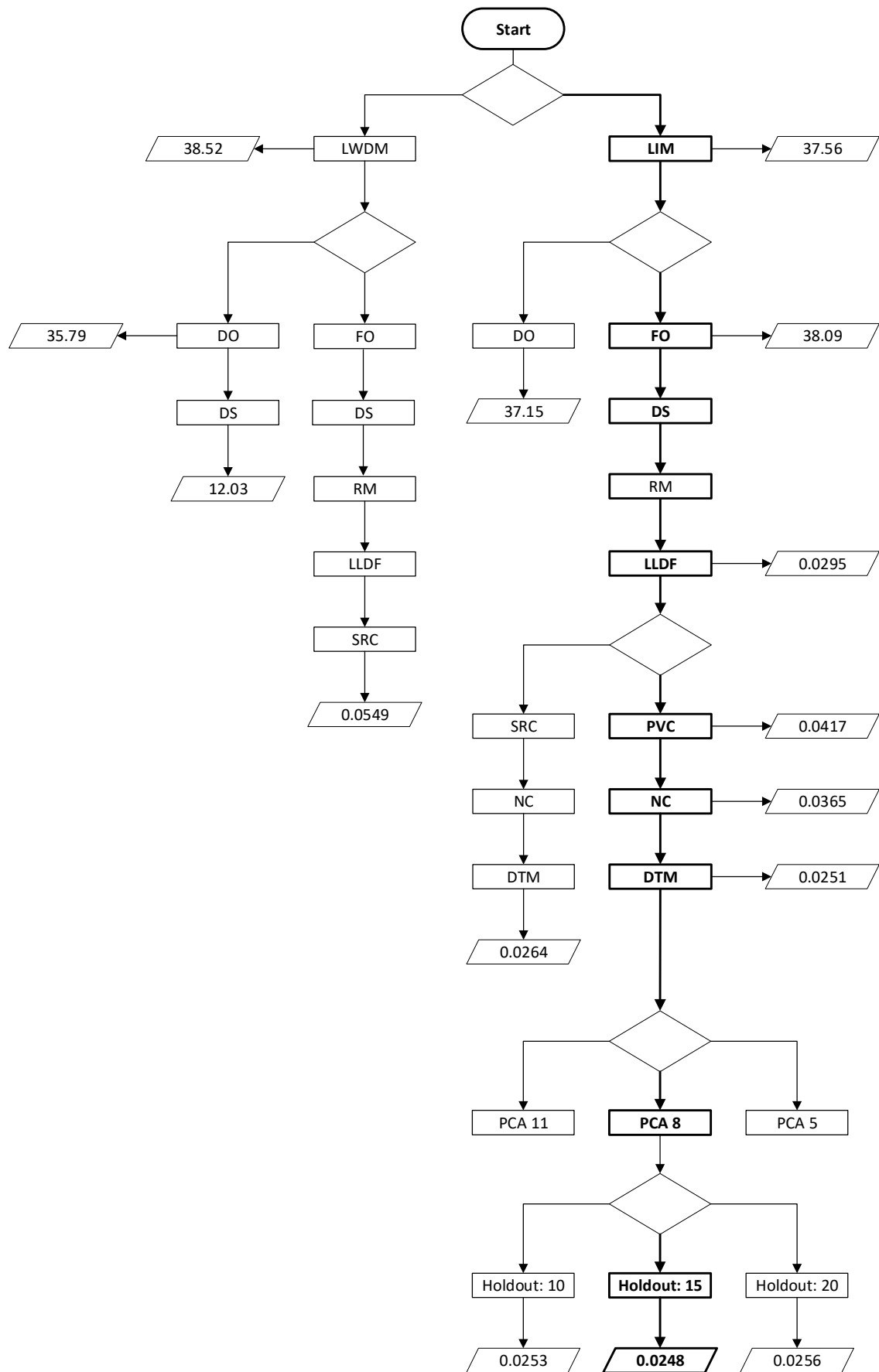


Figure 4-6 Flow chart summary of data preprocessing iterations

Table 4-2 shows a steady improvement of the RMSE from the base-case scenario as each DP method is added. Furthermore, a comparison of the LWDM (RMSE = 38.52) and LIM (RMSE = 37.56) base-case scenarios to the optimal combination of DP methods (RMSE = 0.0248) shows a 100% improvement in the forecasting accuracy. This validates the earlier statement made that a combination of DP methods can significantly increase forecasting accuracy. Finally, the effects of this DP combination is evaluated by comparing the original three-step forecasting architecture (LLDF + LWDM + RM + SRC) with improvements from this study (LLDF + LIM + FO + DS + RM + PVC + NC + DTM+ PCA 8+ Holdout: 15) using the case study data highlighted in Section 4.5.

Table 4-5 shows the results.

Table 4-5 Comparing performance of original Three-step to improved Three-step forecasting model

Original Three-step	Training				Testing			
Model	RMSE	MSE	MAE	R ²	RMSE	MSE	MAE	R ²
MNN	0.0914	0.0084	0.0491	0.86	0.0989	0.0098	0.0531	0.84
Improved Three-step	Training				Testing			
Model	RMSE	MSE	MAE	R ²	RMSE	MSE	MAE	R ²
WNN	0.0254	0.0006	0.0181	0.98	0.0251	0.0006	0.0176	0.98

The DP methods improve the accuracy of the model over the total test sample size by reducing the RMSE error by 75% from 0.0989 to 0.0248. In addition, the 0.98 r-square (R²) value of the improved three-step is greater than 0.84, meaning the predicted results during testing of the improved model are much closer to the actual PV generated than the original three-step model. To visualize the prediction performance, a line graph of the actual vs predicted of both methods over a period of 7 days is shown in Figure 4-7.

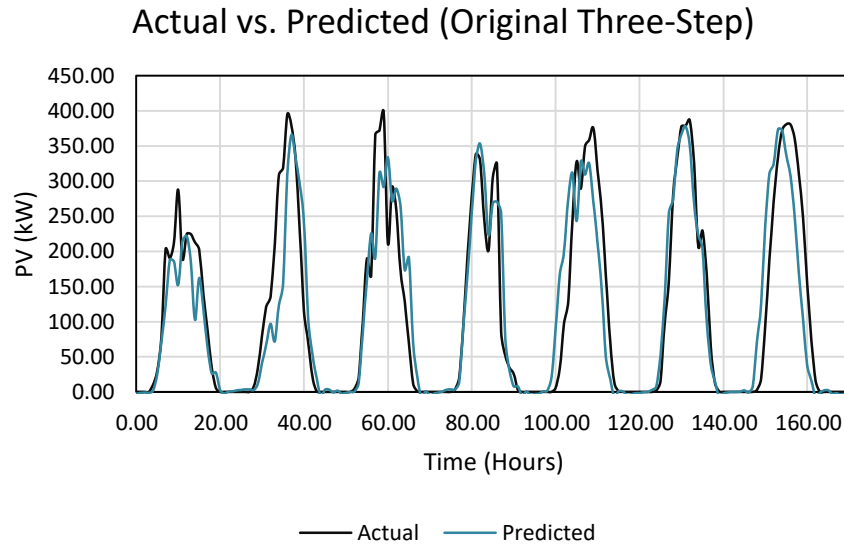
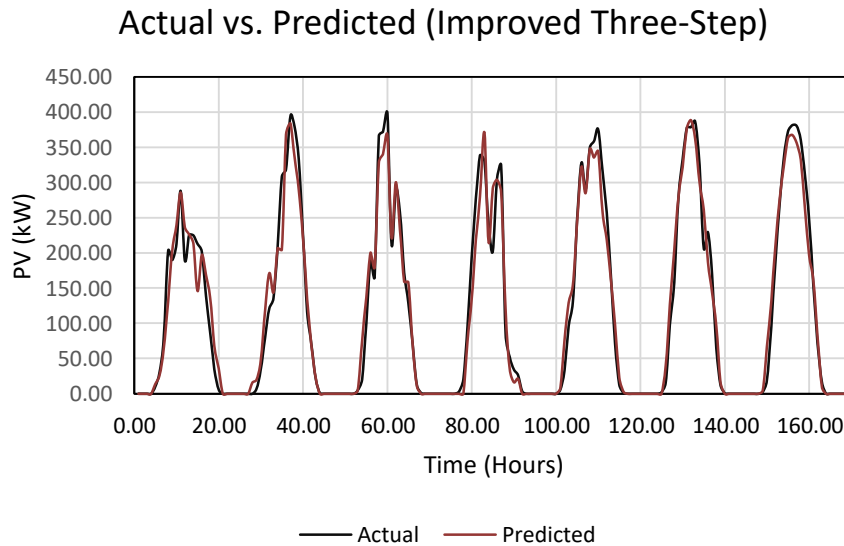


Figure 4-7 One-week prediction comparison between original and improved Three-step forecasting method

The improved three-step model shows better approximation of the actual PV generation, including peaks than the original three-step model which proves the importance of using accurate data in forecasting. In forecasting, hybrid models overcome the limitation of a single technique and ensure better forecasting performance by leveraging on the advantages of each of its input models [89–91]. As has been proven, combining different DP methods also helps to improve forecasting. However, it is also important to note that the performance of different DP methods depends on how the user combines and/or utilizes them, the data type, and the

prediction model. For example, in this study, RM performed better than ZS in all the trials. Arora et al. [92] investigated the various combination of normalization techniques, transfer functions, ANN, and training algorithms, to determine the best methodology for forecasting solar irradiance. The RM, ZS and Decimal score (DSM) normalization techniques were analyzed with DSM yielding the least error when paired with ENN. In another comparison of these three normalization techniques, in a study on feature extraction to identify characters in a Javanese Script [61], raw data normalized by ZS outperformed the other two techniques with 60%-character recognition; better than 56.33% for RM and 29.83% for DS. This means that different normalization techniques tend to work with different forecasting architectures and algorithms. Drawing from this, the outcome of other DP techniques also varies depending on the forecasting architecture. For example, in Appendix B, trials 11 - 16 in Table B-1 show that the same DP methods must be used for both the training and testing data sets otherwise larger RMSE errors will occur. In Table B-2, DP methods like DS and DO were left out while the PPC was adjusted to 0.5 instead of 0.4, leading to higher RMSE values. Table B-3 validates a lot of the findings already established in this study. Table B-4 uses only 1 year data, confirming that the more accurate data used in training the model, the better the prediction.

4.7 Conclusion

This study provides a structured classification of data preprocessing methods for data-driven statistical PV forecasting models. This structured framework will be useful to researchers in assessing which category different preprocessing methods fall under for a more informed and systematic cleaning of data. This study also investigates the impact of sixteen preprocessing methods and their impact on an existing statistical three-step PV forecasting model. Two and a half years of weather and PV generation data were analyzed, with 75% being used for training while the remaining 25% used for testing. The combination of low-level data fusion, linear interpolation, filling outliers, data smoothing, rescaling method, moderate to strong PV correlation of weather parameters using Pearson correlation coefficient, day/time/month decomposition, seasonal decomposition, principal component analysis, and holdout validation improved the accuracy of the three-step forecasting model by reducing the root mean

square error by 75%. Results show that the original three-step PV forecasting model before data preprocessing improvements achieved a root mean square error of 0.0989 and an R^2 value of 0.84 during testing and after data processing achieved 0.0248 and 0.98 respectively. In addition, this combination of data preprocessing methods shows a 100% improvement on forecasting accuracy when compared to a base-case scenario when either the list-wise deletion or linear interpolation method are used proving that combining various preprocessing methods significantly increases model accuracy.

Ultimately, the type of data preprocessing methods used depends on how the user combines and/or utilizes the methods, the data type, and the prediction model. The performance of data preprocessing methods will differ based on the factors mentioned above and the goal of the forecasting problem. Recommendations from this study include testing extensively to see which combination of data preprocessing methods work. In addition, because of overfitting concerns, it is suggested that the resubstitution method be replaced with the cross validation for smaller data samples and holdout validation for larger data sets.

A limitation of the study is that the data used was relatively short. As proven by the differences of the results in the appendix between using one-year data and two and a half years, more data potentially means more accurate data if preprocessed correctly. The data preprocessing methods highlighted in this study are not exhaustive, leaving room for other preprocessing methods to be investigated.

Credit author statement

Tolulope O. Falope: Conceptualization, Methodology, Software, Data curation, Formal Analysis, Visualization, Writing – original draft, **Lao Liyun:** Writing – review & editing, Supervision, **Da Huo:** Writing – review & editing, Supervision.

References

- [1] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews* 2018;81:912–28. <https://doi.org/10.1016/J.RSER.2017.08.017>.
- [2] Pan B, Hirota K, Jia Z, Dai Y. A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods. *Neurocomputing* 2023;561:126866. <https://doi.org/10.1016/J.NEUCOM.2023.126866>.
- [3] Adari GK, Raja M, Vijaya P. Machine learning in genomics: identification and modeling of anticancer peptides. *Data Science for Genomics* 2023;25–68. <https://doi.org/10.1016/B978-0-323-98352-5.00007-0>.
- [4] Liu H, Chen C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl Energy* 2019;249:392–408. <https://doi.org/10.1016/J.APENERGY.2019.04.188>.
- [5] Duong HT, Nguyen-Thi TA. A review: preprocessing techniques and data augmentation for sentiment analysis. *Comput Soc Netw* 2021;8:1–16. <https://doi.org/10.1186/S40649-020-00080-X/FIGURES/4>.
- [6] Jiao Y, Li Z, Chen X, Fei S. Preprocessing methods for near-infrared spectrum calibration. *J Chemom* 2020;34:e3306. <https://doi.org/10.1002/CEM.3306>.
- [7] Albahra S, Gorbett T, Robertson S, D'Aleo G, Kumar SVS, Ockunzzi S, et al. Artificial intelligence and machine learning overview in pathology & laboratory medicine: A general review of data preprocessing and basic supervised concepts. *Semin Diagn Pathol* 2023;40:71–87. <https://doi.org/10.1053/J.SEMDP.2023.02.002>.
- [8] García S, Ramírez-Gallego S, Luengo J, Benítez JM, Herrera F. Big data preprocessing: methods and prospects. *Big Data Analytics* 2016 1:1 2016;1:1–22. <https://doi.org/10.1186/S41044-016-0014-0>.
- [9] Ranjan KG, Prusty BR, Jena D. Review of preprocessing methods for univariate volatile time-series in power system applications. *Electric Power Systems Research* 2021;191:106885. <https://doi.org/10.1016/J.EPSR.2020.106885>.

- [10] Bamm A, Eroshenko S, Khalyasmaa A. Effect of Data Preprocessing on the Forecasting Accuracy of Solar Power Plant. 2021 18th International Scientific Technical Conference Alternating Current Electric Drives, ACED 2021 - Proceedings 2021. <https://doi.org/10.1109/ACED50605.2021.9462288>.
- [11] Phan QT, Wu YK, Phan QD, Lo HY. A Novel Forecasting Model for Solar Power Generation by a Deep Learning Framework With Data Preprocessing and Postprocessing. IEEE Trans Ind Appl 2023;59:220–31. <https://doi.org/10.1109/TIA.2022.3212999>.
- [12] Benhar H, Idri A, L Fernández-Alemán J. Data preprocessing for heart disease classification: A systematic literature review. Comput Methods Programs Biomed 2020;195:105635. <https://doi.org/10.1016/J.CMPB.2020.105635>.
- [13] Pazhoohesh M, Sadegh Javadi M, Gheisari M, Aziz S. Dealing with Missing Data in the Smart Buildings using Innovative Imputation Techniques; Dealing with Missing Data in the Smart Buildings using Innovative Imputation Techniques. IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society 2021. <https://doi.org/10.1109/IECON48115.2021.9612650>.
- [14] Bashir F, Wei HL. Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. Neurocomputing 2018;276:23–30. <https://doi.org/10.1016/J.NEUCOM.2017.03.097>.
- [15] Bashir F, Wei HL. Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm. Neurocomputing 2018;276:23–30. <https://doi.org/10.1016/J.NEUCOM.2017.03.097>.
- [16] Ayilara OF, Tolulope ·, Sajobi T, Barclay · Ruth, Bohm E, Mohammad ·, et al. A comparison of methods to address item non-response when testing for differential item functioning in multidimensional patient-reported outcome measures. Quality of Life Research 2022 2022;1:1–12. <https://doi.org/10.1007/S11136-022-03129-8>.
- [17] Sophia Wang J, Aronow PM. Listwise Deletion in High Dimensions. Political Analysis 2021. <https://doi.org/10.48550/arxiv.2101.11470>.

- [18] Ziyabari S, Du L, Biswas S. Short-term Solar Irradiance Forecasting Based on Multi-Branch Residual Network. ECCE 2020 - IEEE Energy Conversion Congress and Exposition 2020:2000–5. <https://doi.org/10.1109/ECCE44975.2020.9235930>.
- [19] Benchekroun M, Chevallier B, Zalc V, Istrate D, Lenne D, Vera N. The Impact of Missing Data on Heart Rate Variability Features: A Comparative Study of Interpolation Methods for Ambulatory Health Monitoring. IRBM 2023;44:100776. <https://doi.org/10.1016/J.IRBM.2023.100776>.
- [20] Benhar H, Idri A, L Fernández-Alemán J. Data preprocessing for heart disease classification: A systematic literature review. Comput Methods Programs Biomed 2020;195:105635. <https://doi.org/10.1016/J.CMPB.2020.105635>.
- [21] Kim J, Kwak Y, Mun SH, Huh JH. Imputation of missing values in residential building monitored data: Energy consumption, behavior, and environment information. Build Environ 2023;245:110919. <https://doi.org/10.1016/J.BUILDENV.2023.110919>.
- [22] Scott C, Ahsan M, Albarbar A. Machine learning for forecasting a photovoltaic (PV) generation system. Energy 2023;278:127807. <https://doi.org/10.1016/J.ENERGY.2023.127807>.
- [23] Li J, Liu Z. Attribute-weighted outlier detection for mixed data based on parallel mutual information. Expert Syst Appl 2024;236:121304. <https://doi.org/10.1016/J.ESWA.2023.121304>.
- [24] Alimohammadi H, Nancy Chen S. Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis. Expert Syst Appl 2022;191:116371. <https://doi.org/10.1016/J.ESWA.2021.116371>.
- [25] Olszewski D, Iwanowski M, Graniszewski W. Dimensionality reduction for detection of anomalies in the IoT traffic data. Future Generation Computer Systems 2024;151:137–51. <https://doi.org/10.1016/J.FUTURE.2023.09.033>.
- [26] Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. Filter unsupervised spectral feature selection method for mixed data based on a new feature correlation measure. Neurocomputing 2024;571:127111. <https://doi.org/10.1016/J.NEUCOM.2023.127111>.

- [27] Ihianle IK, Machado P, Owa K, Adama DA, Otuka R, Lotfi A. Minimising redundancy, maximising relevance: HRV feature selection for stress classification. *Expert Syst Appl* 2024;239:122490. <https://doi.org/10.1016/J.ESWA.2023.122490>.
- [28] Microsoft. PEARSON function. Microsoft Support 2021. <https://support.microsoft.com/en-us/office/pearson-function-0c3e30fc-e5af-49c4-808a-3ef66e034c18> (accessed October 15, 2021).
- [29] Singla P, Duhan M, Saroha S. A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy* 2021 2021:1–37. <https://doi.org/10.1007/S11708-021-0722-7>.
- [30] Monfared M, Fazeli M, Lewis R, Searle J. Day-Ahead Prediction of PV Generation Using Weather Forecast Data: A Case Study in the UK. 2nd International Conference on Electrical, Communication and Computer Engineering, ICECCE 2020 2020. <https://doi.org/10.1109/ICECCE49384.2020.9179454>.
- [31] Eom H, Son Y, Choi S. Feature-selective ensemble learning-based long-term regional PV generation forecasting. *IEEE Access* 2020;8:54620–30. <https://doi.org/10.1109/ACCESS.2020.2981819>.
- [32] de Jongh S, Riedel T, Mueller F, Yacoub AE, Suriyah M, Leibfried T. Spatio-Temporal Short Term Photovoltaic Generation Forecasting with Uncertainty Estimates using Machine Learning Methods. UPEC 2020 - 2020 55th International Universities Power Engineering Conference, Proceedings 2020. <https://doi.org/10.1109/UPEC49904.2020.9209764>.
- [33] Masood N al, Asif MI, Alam AM, Deeba SR, Aziz T. Forecasting of Photovoltaic Power Generation: Techniques and Key Factors. *Proceedings of 2019 IEEE Region 10 Symposium, TENSYP* 2019 2019:457–61. <https://doi.org/10.1109/TENSYP46218.2019.8971337>.
- [34] Kong J, Oh S, Jung J, Lee I. Improving the accuracy of photovoltaic generation forecasting by considering particulate matter variables. *IEEE Power and Energy Society General Meeting* 2020;2020-August. <https://doi.org/10.1109/PESGM41954.2020.9282054>.

- [35] Zeng J, Wang C, Hu X. Photovoltaic power generation prediction based on correlation analysis and GA - BP neural network. 2021 4th International Conference on Pattern Recognition and Artificial Intelligence, PRAI 2021 2021:306–9. <https://doi.org/10.1109/PRAI53619.2021.9551035>.
- [36] Qin L, Huang W, Guo Z, Zhou Y, Zhang B. Topology Identification Method of Low-voltage Distribution Network Based on Improved Pearson Correlation Coefficient Method 2022:1–6. <https://doi.org/10.1109/CIYCEE53554.2021.9676722>.
- [37] Patel MD, Shah S, Sinha S, Gupta RK, Bharti SK, Singh VP. Solar Irradiation Forecasting - Comparative Analysis of Various Methods. 2021 IEEE International Conference on Technology, Research, and Innovation for Betterment of Society (TRIBES) 2021:1–9. <https://doi.org/10.1109/TRIBES52498.2021.9751626>.
- [38] Xiong Q, Xiong H, Kong Q, Ni X, Li Y, Yuan C. Machine learning-driven seismic failure mode identification of reinforced concrete shear walls based on PCA feature extraction. *Structures* 2022;44:1429–42. <https://doi.org/10.1016/J.ISTRUC.2022.08.089>.
- [39] Kazem HA, Yousif JH, Chaichan MT, Al-Waeli AHA, Sopian K. Long-term power forecasting using FRNN and PCA models for calculating output parameters in solar photovoltaic generation. *Heliyon* 2022;8:e08803. <https://doi.org/10.1016/J.HELİYON.2022.E08803>.
- [40] Lan H, Zhang C, Hong YY, He Y, Wen S. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. *Appl Energy* 2019;247:389–402. <https://doi.org/10.1016/J.APENERGY.2019.04.056>.
- [41] Malvoni M, De Giorgi MG, Congedo PM. Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia* 2017;126:651–8. <https://doi.org/10.1016/J.EGYPRO.2017.08.293>.
- [42] Zhang X, Wei Z. A hybrid model based on principal component analysis, wavelet transform, and extreme learning machine optimized by bat algorithm for daily solar radiation forecasting. *Sustainability (Switzerland)* 2019;11. <https://doi.org/10.3390/SU11154138>.

- [43] Mallika IL, Ratnam DV, Ostuka Y, Sivavaraprasad G, Raman S. Implementation of Hybrid Ionospheric TEC Forecasting Algorithm Using PCA-NN Method. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2019;12:371–81. <https://doi.org/10.1109/JSTARS.2018.2877445>.
- [44] Munawar U, Wang Z. A Framework of Using Machine Learning Approaches for Short-Term Solar Power Forecasting. *Journal of Electrical Engineering & Technology* 2020;15:561–9. <https://doi.org/10.1007/s42835-020-00346-4>.
- [45] Worrell J. *Machine Learning Principal Component Analysis* 2017.
- [46] Ahmad A, Xiao X, Mo H, Dong D. Tuning data preprocessing techniques for improved wind speed prediction. *Energy Reports* 2024;11:287–303. <https://doi.org/10.1016/J.EGYR.2023.11.056>.
- [47] Chen Y, Hu X, Zhang L. A review of ultra-short-term forecasting of wind power based on data decomposition-forecasting technology combination model. *Energy Reports* 2022;8:14200–19. <https://doi.org/10.1016/J.EGYR.2022.10.342>.
- [48] Xie G, Wang S, Lai KK. Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. *J Air Transp Manag* 2014;37:20–6. <https://doi.org/10.1016/J.JAIRTRAMAN.2014.01.009>.
- [49] Fraihat H, Almbaideen AA, Al-Odienat A, Al-Naami B, de Fazio R, Visconti P. Solar Radiation Forecasting by Pearson Correlation Using LSTM Neural Network and ANFIS Method: Application in the West-Central Jordan. *Future Internet* 2022, Vol 14, Page 79 2022;14:79. <https://doi.org/10.3390/FI14030079>.
- [50] Singla P, Duhan M, Saroha S. A Hybrid Solar Irradiance Forecasting Using Full Wavelet Packet Decomposition and Bi-Directional Long Short-Term Memory (BiLSTM). *Arab J Sci Eng* 2022;1–27. <https://doi.org/10.1007/S13369-022-06655-2/TABLES/18>.
- [51] Ali-Ou-Salah H, Oukarfi B, Mouhaydine T. Short-term solar radiation forecasting using a new seasonal clustering technique and artificial neural network. <https://doi.org/10.1080/1543507520211946819> 2021;19:424–34. <https://doi.org/10.1080/15435075.2021.1946819>.

- [52] Kim S-G, Jung J-Y, Sim MK. A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning. *Sustainability* 2019. <https://doi.org/10.3390/su11051501>.
- [53] Shin D, Ha E, Kim T, Kim C. Short-term photovoltaic power generation predicting by input/output structure of weather forecast using deep learning. *Soft Comput* 2021;25:771–83. <https://doi.org/10.1007/S00500-020-05199-7/FIGURES/9>.
- [54] Obiora CN, Hasan AN, Ali A, Alajarmeh N. Forecasting Hourly Solar Radiation Using Artificial Intelligence Techniques. *IEEE Canadian Journal of Electrical and Computer Engineering* 2021;44.
- [55] Rana M, Koprinska I, Agelidis VG. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers Manag* 2016;121:380–90. <https://doi.org/10.1016/J.ENCONMAN.2016.05.025>.
- [56] Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew Energy* 2020;160:26–41. <https://doi.org/10.1016/J.RENENE.2020.05.150>.
- [57] Singh D, Singh B. Investigating the impact of data normalization on classification performance. *Appl Soft Comput* 2020;97:105524. <https://doi.org/10.1016/J.ASOC.2019.105524>.
- [58] Normalize data - MATLAB normalize - MathWorks United Kingdom n.d. <https://uk.mathworks.com/help/matlab/ref/double.normalize.html> (accessed October 30, 2021).
- [59] Chu Y, Li M, Coimbra CFM, Feng D, Wang H. Intra-hour irradiance forecasting techniques for solar power integration: A review. *IScience* 2021;24:103136. <https://doi.org/10.1016/J.ISCI.2021.103136>.
- [60] Santhanakrishnan K, Senthoooran V. A Parallel distributed Cluster Computing Model for Z-Score Computation in Respect of Sri Lankan University Admissions. 2022 2nd International Conference on Advanced Research in Computing (ICARC) 2022:344–8. <https://doi.org/10.1109/ICARC54489.2022.9754093>.
- [61] Herwanto HW, Handayani AN, Wibawa AP, Chandrika KL, Arai K. Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on

- Javanese Character Recognition. 7th International Conference on Electrical, Electronics and Information Engineering: Technological Breakthrough for Greater New Life, ICEEIE 2021 2021. <https://doi.org/10.1109/ICEEIE52663.2021.9616665>.
- [62] Prado-Rujas II, Garcia-Dopico A, Serrano E, Perez MS. A flexible and robust deep learning-based system for solar irradiance forecasting. *IEEE Access* 2021;9:12348–61. <https://doi.org/10.1109/ACCESS.2021.3051839>.
- [63] Herwanto HW, Handayani AN, Wibawa AP, Chandrika KL, Arai K. Comparison of Min-Max, Z-Score and Decimal Scaling Normalization for Zoning Feature Extraction on Javanese Character Recognition. 7th International Conference on Electrical, Electronics and Information Engineering (ICEEE) 2021:1–3. <https://doi.org/10.1109/ICEEIE52663.2021.9616665>.
- [64] Alasadi SA, Bhaya WS. Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences* 2017;12:4102–7. <https://doi.org/10.3923/JEASCI.2017.4102.4107>.
- [65] Hassani S, Dackermann U, Mousavi M, Li J. A systematic review of data fusion techniques for optimized structural health monitoring. *Information Fusion* 2024;103:102136. <https://doi.org/10.1016/J.INFFUS.2023.102136>.
- [66] Munir A, Blasch E, Kwon J, Kong J, Aved A. Artificial Intelligence and Data Fusion at the Edge. *IEEE Aerospace and Electronic Systems Magazine* 2021;36:62–78. <https://doi.org/10.1109/MAES.2020.3043072>.
- [67] Rammelkamp K, Schröder S, Kubitz S, Vogt DS, Frohmann S, Hansen PB, et al. Low-level LIBS and Raman data fusion in the context of in situ Mars exploration. *Journal of Raman Spectroscopy* 2020;51:1682–701. <https://doi.org/10.1002/JRS.5615>.
- [68] Teglia CM, Azcarate SM, Alcaráz MR, Goicoechea HC, Culzoni MJ. Exploiting the synergistic effect of concurrent data signals: Low-level fusion of liquid chromatographic with dual detection data. *Talanta* 2018;186:481–8. <https://doi.org/10.1016/J.TALANTA.2018.04.090>.
- [69] Lanza I, Lolli V, Segato S, Caligiani A, Contiero B, Lotto A, et al. Use of GC–MS and ¹H NMR low-level data fusion as an advanced and comprehensive metabolomic

approach to discriminate milk from dairy chains based on different types of forage. *Int Dairy J* 2021;123:105174. <https://doi.org/10.1016/J.IDAIRYJ.2021.105174>.

- [70] Nguyen QH, Ly HB, Ho LS, Al-Ansari N, Van Le H, Tran VQ, et al. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. *Math Probl Eng* 2021;2021. <https://doi.org/10.1155/2021/4832864>.
- [71] Zhu JJ, Yang M, Ren ZJ. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environ Sci Technol* 2023;57:17671–89. https://doi.org/10.1021/ACS.EST.3C00026/ASSET/IMAGES/LARGE/ES3C00026_0007.JPEG.
- [72] Falope TO, Lao L, Hanak D. A Unique Three-Step Weather Data Approach in Solar Energy Prediction Using Machine Learning. *Sustainable Energy Solutions for a Post-COVID Recovery towards a Better Future: Part VII*, Bangkok: International Conference on Applied Energy ; 2021, p. 1–7. <https://doi.org/https://doi.org/10.46855/energy-proceedings-9863>.
- [73] Pawar P, Mithulananthan N, Raza MQ. Solar PV Power Forecasting Using Modified SVR with Gauss-Newton Method. *Proceedings - 2020 IEEE 2nd Global Power, Energy and Communication Conference, GPECOM 2020* 2020:226–31. <https://doi.org/10.1109/GPECOM49333.2020.9247935>.
- [74] Zhang Z, Li M, Lin M, Soon G, Greene T, Shen C. Subgroup selection in adaptive signature designs of confirmatory clinical trials. *J R Stat Soc Ser C Appl Stat* 2017;66:345–61. <https://doi.org/10.1111/RSSC.12175>.
- [75] Ku Y, Kwon S bin, Yoon J-H, Mun S-K, Chang M. Machine Learning Models for Predicting the Occurrence of Respiratory Diseases Using Climatic and Air-Pollution Factors. *Clin Exp Otorhinolaryngol* 2022. <https://doi.org/10.21053/ceo.2021.01536>.
- [76] Mirza HB, Ratnaparkhe VR. Classifier Tools: A Comparative Study. *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018* 2019:1543–7. <https://doi.org/10.1109/ICCONS.2018.8662904>.
- [77] Ghofrani M, Alolayan M. Time Series and Renewable Energy Forecasting. *Time Series Analysis and Applications* 2017. <https://doi.org/10.5772/INTECHOPEN.70845>.

- [78] Meenal R, Binu · D, Ramya · K C, Prawin ·, Michael A, Vinoth Kumar · K, et al. Weather Forecasting for Renewable Energy System: A Review. *Archives of Computational Methods in Engineering* 2021;1:3. <https://doi.org/10.1007/s11831-021-09695-3>.
- [79] Cranfield University. Cranfield University's Living Laboratory and Urban Observatory - Data. *Living Laboratory* 2021. <https://www.livinglab.ac.uk/data.html> (accessed February 21, 2022).
- [80] Cranfield University. *Energy Manager Live* 2022. <https://cranfield.energymanagerlive.com/> (accessed February 21, 2022).
- [81] Sha'abani MNAH, Fuad N, Jamal N, Ismail MF. kNN and SVM Classification for EEG: A Review. *Lecture Notes in Electrical Engineering* 2020;632:555–65. https://doi.org/10.1007/978-981-15-2317-5_47/TABLES/3.
- [82] Boukraa L, Essahraoui S, Makkaoui K El, Ouahbi I, Esbai R. Intelligent Intrusion Detection in Software-Defined Networking: A Comparative Study of SVM and ANN Models. *Procedia Comput Sci* 2023;224:26–33. <https://doi.org/10.1016/J.PROCS.2023.09.007>.
- [83] National Physical Laboratory. When do the four seasons officially begin? - NPL 2021. <https://www.npl.co.uk/resources/q-a/when-do-the-four-seasons-begin> (accessed May 24, 2022).
- [84] Najibi F, Apostolopoulou D, Alonso E. Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast. *International Journal of Electrical Power & Energy Systems* 2021;130:106916. <https://doi.org/10.1016/J.IJEPES.2021.106916>.
- [85] Zhou Y, Liu Y, Wang D, De G, Li Y, Liu X, et al. A novel combined multi-task learning and Gaussian process regression model for the prediction of multi-timescale and multi-component of solar radiation. *J Clean Prod* 2021;284:124710. <https://doi.org/10.1016/J.JCLEPRO.2020.124710>.
- [86] Sheng H, Xiao J, Cheng Y, Ni Q, Wang S. Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression. *IEEE Transactions on Industrial Electronics* 2018;65:300–8. <https://doi.org/10.1109/TIE.2017.2714127>.

- [87] Sharifzadeh M, Sikinioti-Lock A, Shah N. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. *Renewable and Sustainable Energy Reviews* 2019;108:513–38. <https://doi.org/10.1016/J.RSER.2019.03.040>.
- [88] Lever J, Krzywinski M, Altman N. Points of Significance: Principal component analysis. *Nat Methods* 2017;14:641–2. <https://doi.org/10.1038/NMETH.4346>.
- [89] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews* 2018;81:912–28. <https://doi.org/10.1016/J.RSER.2017.08.017>.
- [90] Shabbir H, Chaturvedi A. Day Ahead Hybrid Forecasting of Global Horizontal Irradiance using Machine Learning (Random Forest Algorithm) and Time-Series Model (SARIMAX). 2022 International Conference on Emerging Smart Computing and Informatics (ESCI) 2022:1–6. <https://doi.org/10.1109/ESCI53509.2022.9758333>.
- [91] Huang CL, Wu YK, Li YY. Deterministic and Probabilistic Solar Power Forecasts: A Review on Forecasting Models. 2021 7th International Conference on Applied System Innovation, ICASI 2021 2021:15–8. <https://doi.org/10.1109/ICASI52993.2021.9568482>.
- [92] Arora I, Gambhir J, Kaur T. Data Normalisation-Based Solar Irradiance Forecasting Using Artificial Neural Networks. *Arab J Sci Eng* 2021;46:1333–43. <https://doi.org/10.1007/S13369-020-05140-Y/FIGURES/4>.

5 Direct Load Control approach: An IoT-based plug load management system for smart buildings⁴

Abstract:

Efficiently managing and controlling electrical consumption and costs in buildings is a crucial objective of today's smart grid and is critical in achieving Net Zero targets. Direct Load Control (DLC) gives utilities, grid operators and customers the capacity to remotely and directly regulate how much power is used by specific end-user devices or loads. While energy management for lighting and heat, ventilation, and air conditioner loads have been addressed with solutions already in the market, small power plug-in loads have not been considered. In this paper, a study on the effect of DLC on managing and controlling small power plug-in load usage at an individual outlet level is carried out using the Baseline comparison method, and by conducting a techno-economic analysis of energy consumption, cost, CO₂ emissions, and payback period. Connected Power, a DLC solution, is used in two sites located in Cranfield University, resulting in potential annual savings of 744.67 kWh, £203.67, 154.22 kg CO₂, for a robotics laboratory and 2993.92 kWh, £818.84, 620.04 kg CO₂ for a café.

Keywords:

Direct load control, Baseline comparison method, Energy management system, Plug-in loads, Demand response

⁴ This chapter is a paper authored by Tolulope Falope, Liyun Lao, & Da Huo, submitted to the Energy journal.

5.1 Introduction

Climate change is one of the world's most urgent problems. According to the Global Carbon Budget [1], fossil fuel contributes about 75% of global greenhouse gas (GHG) emissions and nearly 90% of all CO₂ emissions. A 2022 report by the International Energy Agency (IEA) [2], highlights that the largest increase in emissions came from electricity and heat generation (40%), while buildings accounted for 30% of global final energy consumption, 26% of global GHG emissions and about 11% of the global CO₂ emissions. In terms of energy used in office buildings, plug loads account for between 20% and 50% of total energy consumption used [3–5]. The source of this consumption can be through direct usage, standby or vampire loads. This means that reducing the environmental impact of energy consumption, particularly in buildings, is crucial in solving the climate crisis.

An effective way of reducing energy consumption is by Demand Response (DR). DR is a process whereby the energy demand of a customer is reduced by directly or indirectly shutting off or reducing the consumption of some of the customer's appliances [6]. Direct Load Control (DLC) is one of the most popular and most effective DR programs [7] and is particularly effective in preventing load shedding during critical load periods [8,9]. In DLC programs, system operators can interrupt customer loads during peak or critical hours and in return compensate the customer with a payment or credit [10]. Compensation can be in the form of electricity bill discounts, recurring yearly payments, one-time registration payments, or free hardware installations. DLC leverages advanced communication technologies, such as Internet of Things (IoT) devices and smart meters, smart sockets, smart plugs, smart components (smart thermostats, variable-speed compressors), to control consumer appliances. By strategically limiting or scheduling the energy consumption of these appliances, utilities can reduce grid congestion, prevent blackouts, increase grid reliability, and optimize the use of renewable energy resources.

Apart from cases of peak congestion, DLC can be used to reduce the effect of vampire loads. Vampire loads also referred to as standby power, vampire power, vampire draw, phantom load, ghost load, or leaking electricity, is the amount of electricity used by electrical and electronic appliances when they are turned off (albeit they are made to draw some power) or in standby mode [11]. This definition is also

extended to devices or appliances that draw power when they are not performing their primary function [12,13].

Conventional plug loads can now be referred to as flexible, schedulable, shiftable or interruptible loads through advancement in smart technology. Examples of flexible DLC-controlled loads with high consumption are thermostatically controlled loads (TCLs) like air conditioners, water heaters, and refrigerators [14,15], and other motor-based loads like dishwashers and washing machines. Still other flexible loads are heat pumps, electric water heaters and electric vehicles [16,17], battery banks and devices with accumulators such as drills and vacuum cleaners. On the other side of the spectrum, examples of flexible DLC-controlled loads with small power consumptions are printers, computers, TVs and other household appliances. Building loads are typically categorized into lighting, heating, ventilation, and air conditioner (HVAC) and other small power loads. While lighting and HVAC have solutions to address energy consumption, small power loads remain unaddressed. For example, lighting sensors and timers regulate lighting based on occupancy. To improve HVAC systems, insulation and TCLs are used. There is presently no way to automatically and granularly monitor and/or regulate plug-in/ small power loads.

Connected Power (CP) by Novar ED&S Limited (“Honeywell”) is a load management solution that help saves energy by controlling and managing plug-in power usage, at an individual outlet level. It has the capacity to manage small power loads while integrating them into a building energy management system (BEMS). This paper will investigate how smart sockets, through direct load control and a baseline comparison, can reduce energy consumption and optimise a power system. It will emphasize the role smart sockets play in energy management as well as create a case for its deployment in commercial applications. This paper reviews current DLC frameworks/architectures, identify the gap in addressing small power plug-in loads and propose areas for future research by: (i) analysing the methods of various DLC technologies and categorizing DLC loads, (ii) comparing the architecture of the proposed CP and other DLC systems to identify differences or novelty, and (iii) validating the proposed CP solution in two locations in Cranfield University; a café and a robotics laboratory, to determine how effective the solution is in managing energy consumption. To achieve this, the Baseline comparison method is employed, and a techno-economic analysis is carried out to determine the amount of energy

saved, the saved cost, the equivalent CO₂ emissions reduced, and the payback period of the CP solution.

5.2 Literature review

Moradi-Sarvestani et al. [18] investigated DLC use in reducing distribution system planning costs by 0.5%. In a study by Rama Curiel and Thakur [19], DLC was shown to reduce peak hour demand by up to 2%, eliminating an estimated 117,574 tons of CO₂ and reducing the energy bill of end-users by up to 20% in some cases. Oprea and Bara [20] developed a day-ahead optimization and scheduling with DLC of 114 single-family houses showed electricity bill reduction of up to 22.62% with a 32% improvement of peak to average ratio (PAR). In responding to urgent requests from smart grid operators, Tang et al. [21], showed that by shutting down a calculated number of operating chillers in a building could lead to energy savings of up to 23% of the consumption contributed by the chillers. In total, about 34% of the total power demand could be reduced. In a similar study of AC units, Wang and Tang [22] propose a novel cooling supply-based feedback control strategy with DLC to deal with the power imbalance of a fast demand request. The proposed DR strategy could achieve a power reduction of about 24% while maintaining acceptable thermal comfort levels. In yet another study on AC systems, Dai et al [23] proposed multi-agent based distributed cooperative control saved a total of 2,562 kWh or 19.7% of energy consumption using conventional control, and also reduces power rebound effect by 1,360 kW or 16.5% of power consumption. Hlalele et al. [27] consider participants with home energy management system that include DLC controlled electric water heaters. The load is modelled on a modified IEEE 30-bus (6 thermal units and 41 transmission lines) and IEEE 118-bus (54 thermal generators, 186 transmission lines with 5 PV and wind farms each) test system. DLC integrated into an energy management system (EMS) was proposed by Erenoglu et al. [15] to reduce losses in the mini-grid by scheduling operational periods of Acs and energy storage systems (ESS), achieving about 2.65% in loss reduction. In a simulation of 200 residential houses with smart wifi AC thermostats, Adhikari et al. [24] showed that DLC reduced peak power by up to 60% while maintaining preset temperature levels. Hosseini et al. [8] study of multi-carrier energy systems shows that DLC reduced load shedding of high priority loads by up to 66%. To improve power reductions, Oprea and Bara proposes an adaptive

DLC that integrates day-ahead optimization using fog and edge computing. In an analysis of 114 single-family houses, the proposed DLC architecture reduced the number of interruptions by 20.51% about and improved PAR by 32%. Table 5-1 shows a summary of some DLC studies with respect to the end-user load.

Table 5-1 Summary of direct load control studies and end-user load

Study	Load	Demand strategy	Objective	System
Liu et al. [9]	AC	DLC + TOU	Peak load and peak-valley difference (PVD) reduction	
Moradi-Sarvestani et al. [18]	Feeder	DLC + DDG	Peak to average load ratio (PAR) reduction	DSP
Rama Curiel and Thakur [19]	AC	DLC	Peak load reduction	
Oprea and Bara [20]	pool pump, water heater, AC, electric heating and so on	DLC + IoT + MQTT	Electricity cost reduction	
Tang et al. [21]	AC (chillers/pumps)	DLC	Maintain thermal comfort threshold, reduce building cooling and power demand predictors	Commercial buildings in smart grids
Van Putten et al. [14]	(VOCACs)	DLC	Power reduction	
Neglia et al. [25]	Electric appliances	DLC + teletraffic tools	Peak shaving	
Zhang and de Dear [26]	AC	DLC + Taguchi method	Maintain thermal comfort and cognitive performance outcome	Commercial buildings
Wang and Tang [22]	AC	DLC + cooling supply-based feedback control	Power reduction	Commercial buildings
Dai et al. [23]	AC	DLC + Multi-agent based distributed cooperative control	Even thermal comfort	Commercial buildings
Hlalele et al. [27]	EWHs	DLC + Renewable obligation policy + TOU	Minimise generation cost and maximize RES penetration	Sub-stations
Erenoglu et al. [15]	ACs, ESS	DLC + EMS	Maximising RES penetration, minimise losses, remain within comfort and operational constraints	Mini-grid
Adhikari et al. [24]	HVAC	DLC	Minimise aggregated demand	Residential houses

Prathapaneni and Detroja [28]	Reverse osmosis desalination plant	DLC	Coordinated sizing of microgrid with reverse osmosis desalination plant while minimising system lifetime cost	Reverse osmosis desalination plant
Hosseini et al. [8]	Boiler, electric heat pump, absorption chiller, Thermal electrical and cooling storage systems	DLC + ESS + P2G	Minimise total operating costs	Residential, industrial and commercial electricity-gas-heating networks
Mohseni et al. [7]	EVs and residential appliances	DLC + moth-flame optimization algorithm	Optimal sizing	Off-grid micro-grids
Lankeshwara et al. [29]	ACs, EWHs	DLC + control algorithm	Demand reduction	Residential loads
Jabian et al. [30]	Smart customer appliances	DLC + Genetic algorithm	Automate load shedding	Smart distribution system
Oprea and Bara [31]	Shiftable, interruptible, and battery loads	DLC + Edge and Fog computing	Reduce number of interruptions, Peak to average load ratio (PAR) reduction	Residential loads
This study	Small power plug-in load	DLC	Reduce energy consumption	Commercial buildings

Note: AC = Air conditioner(s), DDG = Decentralized distribution generation, DSP = Distribution system planning, EWH = Electric water heater(s), EMS = Energy Management System, ESS = Energy storage system, MQTT = Message Queueing Telemetry Transport, P2G = Power to gas, RES = Renewable energy sources, VOCACs = Variable-output compressor air conditioners

From the summary in Table 5-1, the “Load” column describes DLC appliances involved in the study. It is clear to see that there is a lack of field data for small power plug-in load. The electrical appliances mentioned in the table refer to simulated plug-in loads, creating a need for validation with experimental live data from real-life appliances.

5.3 Methods

5.3.1 System description

CP is made up of smart sockets and the hub. Smart sockets were developed to enable remote communication and control of household electric equipment that plug into wall sockets [32]. Smart sockets provide control to non-smart loads making them part of the smart system or grid. It is usually installed between a home appliance and the power outlet, and exchanges information with an intelligent interactive terminal also known as the hub, control centre, gateway or controller [33]. Smart sockets will perform the actual load control based on instructions received from the hub. The

communication between the hub and smart sockets is bidirectional. One hub can support up to 50 sockets while the system allows for up to 50 hubs to run together. This gives a system capacity of 2500 sockets or 5000 individual outlets. As can be seen in Figure 5-1, the hub communicates with the smart socket via an innovative wireless network called Mira Mesh RF from Lumen Radio. This is then interlinked with the buildings BEMS via Building Automation and Control Network over Internet Protocol (BACnet/IP).

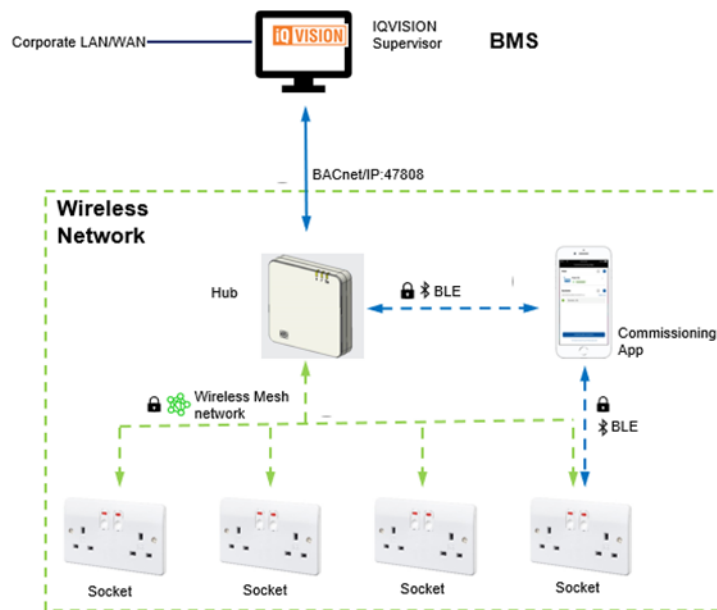


Figure 5-1 Network architecture for Connected Power

Further analysis of DLC studies reveal that CP's architecture is the only solution that has the unique combination of Mira Mesh RF and BACnet/IP integrated into a BEMS. Table 5-2 shows a summary of these studies.

Table 5-2. Studies showing direct load control architecture

Study	Database	Web service	Features	Hardware
Phangbertha et al. [32]	MySQL 5.7.26	Apache/2.4.18 (Ubuntu)	<ul style="list-style-type: none"> * Current monitoring * Switching on/off the socket * Set timer/scheduling 	<ul style="list-style-type: none"> * Wemos D1 Mini for module development board * Relay 2 Channel for electricity switch * Power Supply Module to convert 220V AC to 5V DC * Electricity Sensor to read AC/DC current * Analog to Digital Sensor

Ahmed et al. [34]	Personal computer	ZigBee wireless communication	* Current and voltage monitoring * Switching on/off the socket * Temperature measurement	Zigbee modules
Pawar and Vittal K [35],		ZigBee (802.15.4)	* RMS voltage and current monitoring * Computes apparent power, real power and energy * Switching on/off the socket	* Zigbee series-2 module * LEM LV-25P Voltage sensor * Relay module * ATMEGA328 MCU * LEM LA-55P Current sensor
Zou et al. [36]	Aqara data-stored cloud platform	ZigBee wireless communication	Power monitoring	Aqara smart socket
Chiu et al. [37]	Cloud platform	802.11 b/ g/n Wi-Fi network	Instant power, current/Amperage, voltage/Volts, and the cumulative power consumption in kilowatt hours for the day, week, month, and year.	Smart plug socket
Neglia et al. [25]	Cloud via Data concentrator	Zigbee or PLC	Peak shaving	Smart plug
This study	Cloud	* Mira Mesh RF * BACnet/IP	* Load scheduling: switching on/off the socket * Temperature monitoring	* Logic plus smart sockets by MK electric * Hub * Computer running IQ Vision Supervisor

5.3.2 Baseline comparison method

The baseline comparison method is used to determine the amount of energy saved by the CP solution. Firstly, it establishes a baseline for energy consumption without energy-saving measures. After the energy-saving measures have been implemented, the newly established baseline is then compared to actual consumption to quantify savings. Ocampo Batlle et al. [38], employs a methodology for creating energy baselines in higher education institution buildings to identify areas of potential electrical energy consumption savings. He et al. [39] uses a similar baseline concept to estimate energy savings from improvements in occupant behaviour. In this study, the baseline comparison method is broken down into three phases: baseline, scheduling, and optimisation. The baseline observes the consumption of devices and

identifies peak/off-peak times. Peak times are times when the devices are likely to be in use (normal hours of operation) while off-peak times are times when devices can be shut-off. This time segmentation is important because each time-period will have its own unique load profile defined by the customer's load consumption and would require its own energy saving strategy. The baseline of the customer's load consumption makes up their historical load profile.

Once the baseline has been identified, the controller turns-on and turns-off the loads via the smart sockets during peak and off-peak times respectively. This is known as the scheduling phase. The objective of this phase is to reduce or eliminate vampire loads and device idling times by scheduling appliances to only come on during times when they are needed and have historically operated.

The last phase is the optimization phase. Here, recommendations by the end-user and the output from the scheduling phase can be used to identify potential savings opportunities in both peak and off-peak time periods. For example, additional savings may be realized during peak times by requesting devices to be manually turned-on rather than automatically scheduling them, if these devices display inconsistent consumption patterns. In this way, the hours of operation of idling or devices on standby are reduced, shaving peak hours and adding them to off-peak hours. Figure 5-2 shows the sequence of the baseline comparison method.

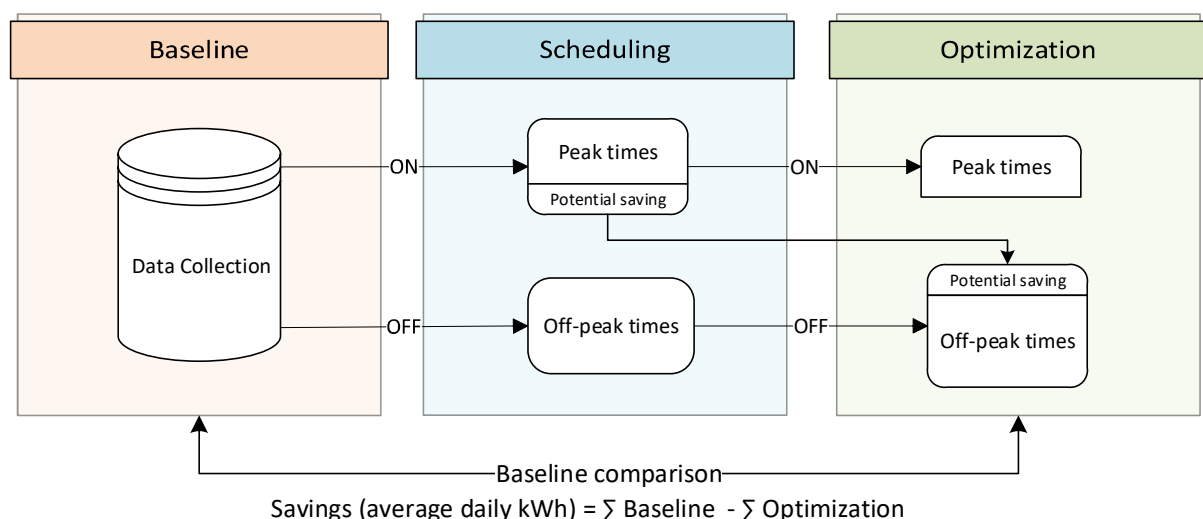


Figure 5-2 Baseline comparison method for estimating energy savings

5.3.3 CO₂ emissions

The estimated CO₂ emissions is calculated by multiplying the energy saved and the UK electricity conversion factor as shown in Equation (5-1) [40]. The national average for UK Scope 2 electricity emissions as recommended by the Department for Energy Security and Net Zero is equivalent to 0.2071 kgCO₂e.

$$CO_2 \text{ Emissions} = \text{Energy saved (kwh)} \times 0.2071 \text{ (kgCO}_2\text{e)} \quad (5-1)$$

5.3.4 Payback period

The payback period (PP) is the length of time required to recover the initial investments cost [41].

$$PP = \frac{\text{Capital costs}}{\text{Annual savings}} \quad (5-2)$$

To determine annual savings, the electricity price must be known. According to The Office of Gas and Electricity Markets (Ofgem), the energy price cap between 1st October to 31st December 2023, which coincides with the duration of this experiment, was 27.35 pence per kWh [42].

5.3.5 Case study

Two sites in Cranfield University were selected for the case study: a café and a robotics laboratory. For each site, five double pre-wired MK Electric smart sockets, a HUB and a PC running the IQ Vision Supervisor were provided. A summary of the assigned sockets and outlets are given in Table 5-3 Socket and outlet designations.

Table 5-3 Socket and outlet designations

Socket	Outlet	Robotics laboratory	café
1	1	3D printer	Panini machine
	2	Network router	Laptop
2	3	ABB robot	Dish washer
	4	Vacant	Milk fridge
3	5	PC/Monitor/Desktop	Fridge bulk brew
	6	PLC	Vacant
4	7	2x monitors	Backbar chiller 1
	8	PC	Backbar chiller 2
5	9	Mobile battery charger	Upright display chiller 1
	10	Robot dog	Upright display chiller 2

The choice of small power loads was selected by managers of each site and connected to the smart sockets as can be seen in Figure 5-3. The analysis at both locations were run for a period of 70 days, between August and October 2023. Both locations represent differing load profiles. The café uses more energy and has a more varied demand due to customer numbers while the laboratory is more constant and defined, representing more of a typical office space.

Café



Single and double upright display chillers

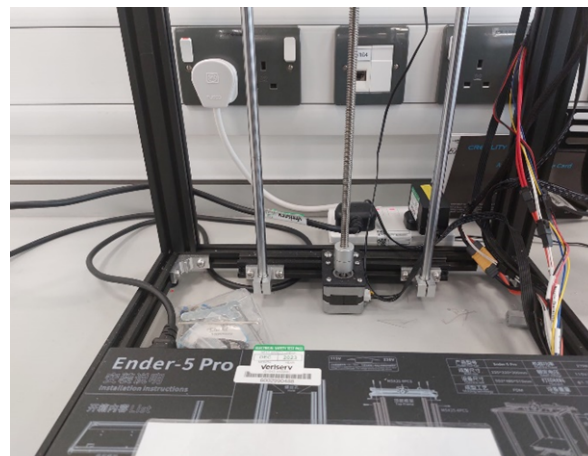


Backbar chiller

Robotics laboratory



PC/Monitor/Desktop



3D Printer

Figure 5-3 Small power plug-in loads connected to MK Electric smart sockets

5.4 Results and discussion

For the robotics laboratory, the phase timelines were split into 16 days/25 days/ 29 days for the baseline/scheduling/optimization respectively. The working hours of the laboratory are between 8am to 6pm. These working hours are assigned to peak times while 6pm to 8am are assigned to off-peak times. Figure 5-4 shows the power consumed for each half-hour on an arbitrary day in August.

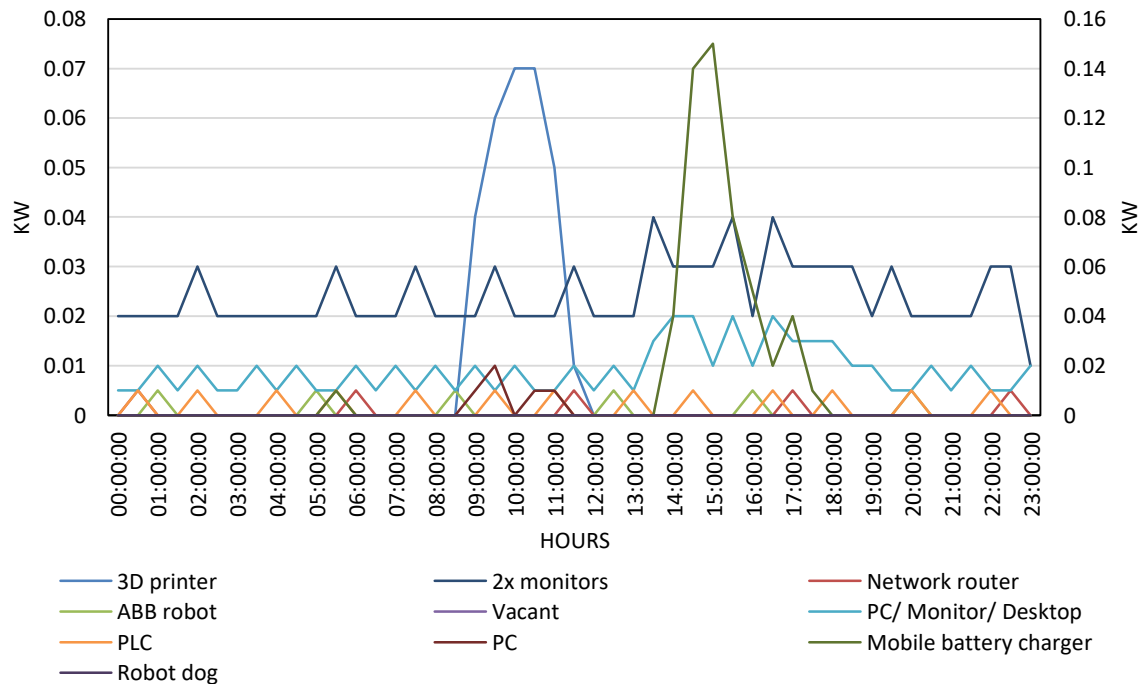


Figure 5-4 Robotics laboratory half-hourly load consumption - 01/08/2023

With the device patterns identified and split into peak and off-peak times, the outlets are scheduled to always turn off from 6pm to 8am during weekdays and during weekends. This schedule is maintained for 25 days. During the optimization phase and with input from the lab manager and the load profile, we notice that not all devices are turned on at 8am. In fact, some may not be turned on at all while others may be turned on later depending on the staff present. It is therefore proposed that beyond turning the devices off between 5pm to 8am, the devices are left off until when they are needed. The CP MK electric smart sockets can operate like a normal socket while still maintaining a programmed schedule. The optimization phase is run for 29 days.

For the café, the phase timelines are split into 14 days/15 days/ 41 days for the baseline/scheduling/optimization respectively. The randomness of the split was due to interruptions of the experiment in the different phases. The operating hours of the café

are from 8:30am to 4:30pm. However, the staff get in at 8:15am and advised that the equipment be turned on at 8am. The appliances were then scheduled to operate from 8am to 5pm. During the optimization phase, it was recommended that the closing time be pushed up to 4:30pm as business had largely been closed for the day. Due to the fears of some perishables spoiling, it was also recommended that one chiller be left on in case of equipment failure. One point to note is that in both locations, potential savings were realized by reducing the peak times and adding them to the off-peak times as stated in Figure 5-2. The output of the optimization is then compared to the baseline to realize the potential savings. To get a more granular look at the data, the results are separated by device/appliance, based on peak/off-peak times and weekdays/weekends. The results of the trial in both locations are shown in Table 5-4.

Table 5-4 Results of the Baseline comparison method

Robotics Laboratory													
Socket	Outlet	Device/ Appliance	Days	Phase	Weekdays (kWh)		Weekends (kWh)		Weekday Total (kWh)	Weekend Total (kWh)	Total (kWh)	Avg. Daily Total (kWh)	% diff respect to B
					8:00 – 18:00	18:00 – 8:00	8:00 – 18:00	18:00 – 8:00					
1	1	3D printer	16	B	2.62	0	0	0	2.62	0	2.62	0.16	-
			25	S	3.22	0	0	0	3.22	0	3.22	0.13	-21%
			29	O	5.38	0	0	0	5.38	0	5.38	0.19	13%
	2	Network router	16	B	2.93	0.23	0.07	0.11	3.16	0.18	3.34	0.21	-
			25	S	0.59	0	0	0	0.59	0	0.59	0.02	-89%
			29	O	1.83	0.06	0	0	1.89	0	1.89	0.07	-69%
2	3	ABB robot	16	B	0.28	0.36	0.12	0.13	0.64	0.25	0.89	0.06	-
			25	S	0.28	0	0	0	0.28	0	0.28	0.01	-80%
			29	O	0	0	0	0	0	0	0	0.00	-100%
	4	Vacant	16	B	0	0	1.35	1.74	0	3.09	3.09	0.19	-
			25	S	0.19	0.43	0	0	0.62	0	0.62	0.02	-87%
			29	O	3.04	0.88	0	0	3.92	0	3.92	0.14	-30%
3	5	PC/Monitor /Desktop	16	B	3.34	3.65	1.17	1.5	6.99	2.67	9.66	0.60	-
			25	S	6.06	0.03	0	0	6.09	0	6.09	0.24	-60%
			29	O	0	0	0	0	0	0	0	0.00	-100%
	6	PLC	16	B	0.59	0.72	0.24	0.31	1.31	0.55	1.86	0.12	-
			25	S	1.58	0.02	0	0	1.6	0	1.6	0.06	-45%
			29	O	0.59	0.44	0	0	1.03	0	1.03	0.04	-69%
4	7	2x monitors	16	B	5.24	6.24	2.05	2.62	11.48	4.67	16.15	1.01	-
			25	S	6.5	0.02	0	0	6.52	0	6.52	0.26	-74%
			29	O	8.13	0	0	0	8.13	0	8.13	0.28	-72%
	8	PC	16	B	0.5	0	0	0	0.5	0	0.5	0.03	-
			25	S	2.5	0	0	0	2.5	0	2.5	0.10	220%

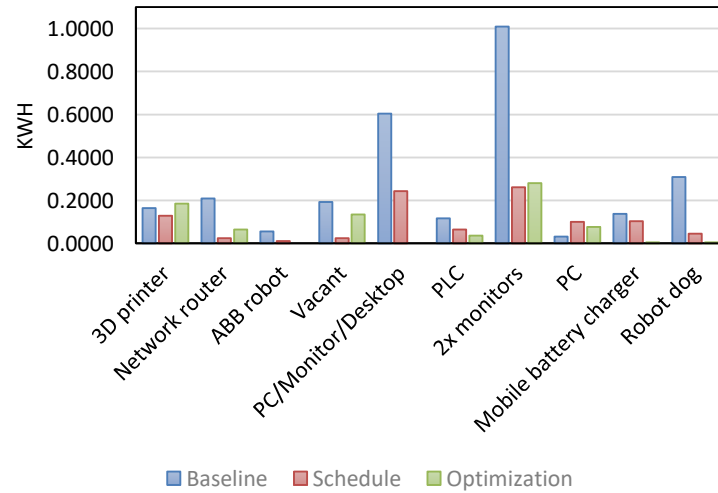
			29	O	2.21	0	0	0	2.21	0	2.21	0.08	144%
5	9	Mobile battery charger	16	B	1.62	0.58	0	0	2.2	0	2.2	0.14	-
			25	S	1.26	0.73	0.25	0.33	1.99	0.58	2.57	0.10	-25%
			29	O	0.15	0	0	0	0.15	0	0.15	0.01	-96%
	10	Robot dog	16	B	3.97	0.97	0	0	4.94	0	4.94	0.31	-
			25	S	1.14	0	0	0	1.14	0	1.14	0.05	-85%
			29	O	0.14	0	0	0	0.14	0	0.14	0.00	-98%
Total Baseline			16	B	21.09	12.75	5	6.41	33.84	11.41	45.25	2.83	
Total Scheduling			25	S	23.32	1.23	0.25	0.33	24.55	0.58	25.13	1.01	-64%
Total Optimization			29	O	21.47	1.38	0	0	22.85	0	22.85	0.79	-72%
Average Daily Savings (kWh)												2.04	
Café													
Socket	Outlet	Device/ Appliance	Days	Phase	Weekdays (kWh)		Weekends (kWh)		Weekday Total (kWh)	Weekend Total (kWh)	Total (kWh)	Avg. Daily Total (kWh)	% diff respect to B
					8:00 – 18:00	18:00 – 8:00	8:00 – 18:00	18:00 – 8:00					
1	1	Panini machine	14	B	43.3	0.07	0	0	43.37	0	43.37	3.10	-
			15	S	39.71	0	0	0	39.71	0	39.71	2.65	-15%
			41	O	61.23	0.06	0	0	61.29	0	61.29	1.49	-52%
	2	Laptop	14	B	0.06	0	0	0	0.06	0	0.06	0.00	-
			15	S	0.04	0	0	0	0.04	0	0.04	0.00	-38%
			41	O	0.09	0	0	0	0.09	0	0.09	0.00	-49%
2	3	Dish washer	14	B	0.15	0.27	0.07	0.09	0.42	0.16	0.58	0.04	-
			15	S	0.16	0.01	0.06	0	0.17	0.06	0.23	0.02	-63%
			41	O	0.39	0	0.14	0	0.39	0.14	0.53	0.01	-69%
	4	Milk fridge	14	B	21.64	31.98	8.82	13.58	53.62	22.4	76.02	5.43	-
			15	S	13.82	3.37	3.74	0	17.19	3.74	20.93	1.40	-74%
			41	O	65.95	85.54	25.73	34.14	151.49	59.87	211.36	5.16	-5%
3	5		14	B	21.36	31.1	8.72	13.38	52.46	22.1	74.56	5.33	-

		Fridge bulk brew	15	S	21.36	1.61	7.63	0	22.97	7.63	30.6	2.04	-62%
			41	O	54.9	0	17.68	0	54.9	17.68	72.58	1.77	-67%
	6	Vacant	14	B	0	0	0	0	0	0	0	0.00	-
			15	S	0	0	0	0	0	0	0	0.00	0%
			41	O	0	0	0	0	0	0	0	0.00	0%
4	7	Backbar chiller 1	14	B	3.11	4.15	1.13	1.77	7.26	2.9	10.16	0.73	-
			15	S	3.18	0.23	1.02	0	3.41	1.02	4.43	0.30	-59%
			41	O	56.38	0.46	0.93	0	56.84	0.93	57.77	1.41	94%
	8	Backbar chiller 2	14	B	3.26	4.19	1.19	1.77	7.45	2.96	10.41	0.74	-
			15	S	3.3	0.23	1.05	0	3.53	1.05	4.58	0.31	-59%
			41	O	3.72	0.45	0.93	0	4.17	0.93	5.1	0.12	-83%
5	9	Upright display chiller 1	14	B	9.08	4.17	1.27	1.72	13.25	2.99	16.24	1.16	-
			15	S	11.05	0.22	2.11	0	11.27	2.11	13.38	0.89	-23%
			41	O	12.62	0.48	1.6	0	13.1	1.6	14.7	0.36	-69%
	10	Upright display chiller 2	14	B	12.95	12.16	2.88	4.32	25.11	7.2	32.31	2.31	-
			15	S	9.93	0.63	2.57	0	10.56	2.57	13.13	0.88	-62%
			41	O	9.57	1.09	1.91	0	10.66	1.91	12.57	0.31	-87%
Total Baseline			14	B	114.91	88.09	24.08	36.63	203	60.71	263.71	18.84	-
Total Scheduling			15	S	102.55	6.3	18.18	0	108.85	18.18	127.03	8.47	-55%
Total Optimization			41	O	264.85	88.08	48.92	34.14	352.93	83.06	435.99	10.63	-44%
Average Daily Savings (kWh)												8.20	

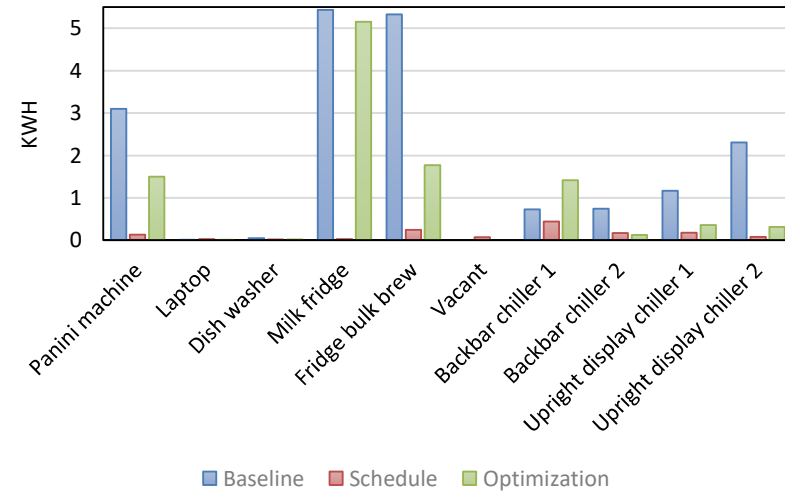
* B = Baseline, S = Schedule, O = Optimization

In addition, the data points in Table 5-4 are plotted in two sets of graphs in Figure 5. The first graphs show the average daily kWh over the total number of days in each phase. The second graphs compare the weekday and weekend consumption of each device/appliance. By identifying the biggest energy user and consumption pattern, the user is better informed of the area with the most potential saving.

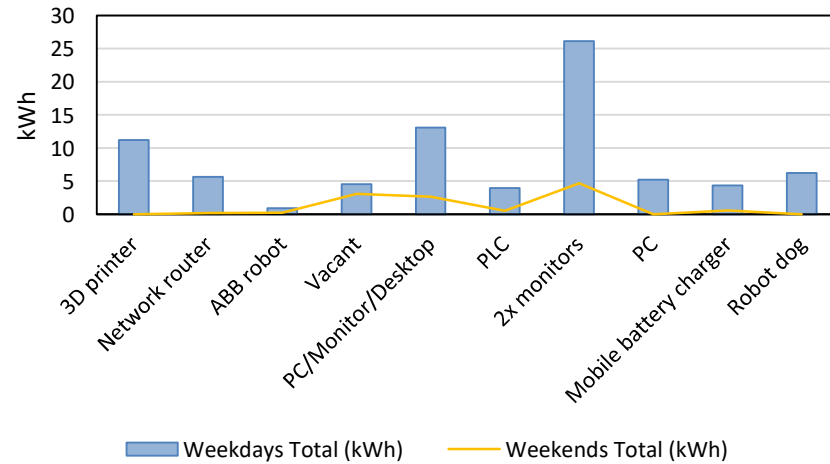
(a) Robotics laboratory – Average daily use



(b) Café – Average daily use



(c) Robotics laboratory - Weekday vs Weekend



(d) Cafe - Weekday vs Weekend

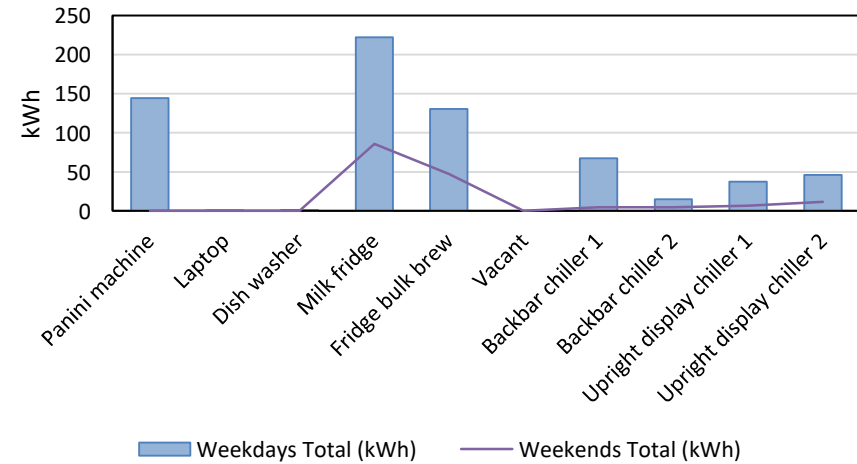


Figure 5-5 Device consumption comparison in kWh

In Table 5-4, the consumption of each device and appliance during each phase is shown, culminating in a daily average consumption. The biggest energy consumers in the robotics laboratory are the 2x monitors and PC/Monitor/Desktop. It was confirmed that the user permanently leaves them on 24 hrs a day even when not in use. For the café, the biggest energy users are the milk fridge and the fridge bulk brew. To get the average daily savings, the average daily consumption of each device/appliance during the baseline phase is summed up and subtracted from the average daily consumption during the optimization phase as suggested in Figure 5-2. This is equivalent to 2.04 kWh and 8.2 kWh for the robotics laboratory and café respectively. While optimization generally brings down the baseline consumption, that is not always the case as can be seen in the 3D printer and backbar chiller in Figures 5-6a and 5-6b. Interestingly in Figure 5-6b, the optimization and baseline are very high for the milk fridge, while the consumption during the scheduling phase is negligible. This is because the optimization is sometimes influenced by the decision of the end user. In this case, the end user did not want the milk fridge to be scheduled off at any time due to fears of products going bad. While the fridge was turned off during the scheduling phase, the original 24 hr schedule was re-instated during the optimization phase. Figure 5-6c and 5-6d show the biggest users during the weekday and weekend and support the findings from 5-6a and 5-6b. As expected, the weekday use is more than weekend for every device in a commercial/office setting. However, weekend and off-peak trends need to be monitored as they can contribute significantly to the consumption as shown by the yellow line in the graph.

Finally, the payback period and annual CO₂ emissions are calculated from the savings highlighted in Table 5-4 using Equations (5-2) and (5-1) respectively. The capital cost of the CP solution (five double pre-wired MK Electric smart sockets, a HUB and the IQ Vision Supervisor license) for each location is £1686.08. The annual saving is determined by multiplying the electricity price of 27.35p by the annual energy saving of each device and summing them up.

Table 5-5 Results of techno-economic analysis

Robotics Laboratory				
Device/ Appliance	kWh saving per day	kWh saving per year	Annual cost saving	Annual CO ₂ saving
3D printer	-0.02	-7.95	-£2.17	-1.65
Network router	0.14	52.41	£14.33	10.85
ABB robot	0.06	20.30	£5.55	4.20
Vacant	0.06	21.15	£5.79	4.38
PC/Monitor/Desktop	0.60	220.37	£60.27	45.64
PLC	0.08	29.47	£8.06	6.10
2x monitors	0.73	266.10	£72.78	55.11
PC	-0.04	-16.41	-£4.49	-3.40
Mobile battery charger	0.13	48.30	£13.21	10.00
Robot dog	0.30	110.93	£30.34	22.97
Total savings	2.04	744.67	£203.67	154.22
Capital costs		Annual cost savings		Payback period
£1686.08		£203.67		8.28 years
Café				
Device/ Appliance	kWh saving per day	kWh saving per year	Annual cost saving	Annual CO ₂ saving
Panini machine	1.60	585.09	£160.02	121.17
Laptop	0.00	0.76	£0.21	0.16
Dish washer	0.03	10.40	£2.85	2.15
Milk fridge	0.27	100.33	£27.44	20.78
Fridge bulk brew	3.56	1297.75	£354.93	268.76
Vacant	0.00	0.00	£0.00	0.00
Backbar chiller 1	-0.68	-249.41	-£68.21	-51.65
Backbar chiller 2	0.62	226.00	£61.81	46.80
Upright display chiller 1	0.80	292.53	£80.01	60.58
Upright display chiller 2	2.00	730.46	£199.78	151.28
Total savings	8.20	2993.92	£818.84	620.04
Capital costs		Annual cost savings		Payback period
£1686.08		£818.84		2.06 years

Table 5-5 shows the robotics laboratory has a potential annual saving of 744.67 kWh, £203.67, 154.22 kg CO₂, and a payback period of 8.28 years, while the café has a potential annual saving of 2993.92 kWh, £818.84, 620.04 kg CO₂ with a payback period of 2.06 years. This corresponds to savings of 44% and 72% when compared to the robotics laboratory and café baselines respectively. Intuitively, the CP solution will create higher savings in sites with potentially higher vampire loads, accounting for the vast difference in annual savings and payback periods between the two sites. Therefore, the number of connected appliances, load type and load usage patterns are critical in site selection. Note that the negative numbers in the kWh and savings columns simply imply that the consumption during the optimization period window was more than that of the baseline. This could be due to a change in consumption patterns, with more information needed to clarify the discrepancy. In any case, the CP solution has the potential to generate great benefits for both the user and the environment.

Conclusion

Direct load control is a crucial demand response and can help reduce waste and consumption when properly implemented. Connected Power by Honeywell is a load management solution that uses a unique combination of MK electric smart sockets, a gateway controller, and the Mira Mesh RF wireless network, to save energy by controlling and managing plug-in power usage, at an individual outlet level. By using the Baseline comparison method, Connected Power reduced energy consumption by 44% and 72% in a robotics laboratory and café respectively. A techno-economic analysis was performed resulting in potential annual savings of 744.67 kWh, £203.67, 154.22 kg CO₂, for the robotics laboratory and 2993.92 kWh, £818.84, 620.04 kg CO₂ for the café.

The trial in each location lasted for 70 days across three different phases: baseline, schedule, and optimization. A recommendation for future research is to increase the number of days in each phase to better identify usage patterns and reduce the standard deviation due to the sporadic nature of the power usage data. In addition, trials can be held throughout the year to investigate the impact

of seasonality. Other drivers of energy consumption like occupancy and holidays can be captured to have a more unbiased assessment of the average energy used. Lastly, the trials can be extended to residential accommodation, and a savings comparison made between residential and commercial buildings.

Credit author statement

Tolulope O. Falope: Conceptualization, Methodology, Software, Data curation, Formal Analysis, Visualization, Writing – original draft, **Lao Liyun:** Writing – review & editing, Supervision, **Da Huo:** Writing – review & editing, Supervision.

References

- [1] Friedlingstein P, O'sullivan M, Jones MW, Andrew RM, Gregor L, Hauck J, et al. Global Carbon Budget 2022. *Earth Syst Sci Data* 2022;14:4811–900. <https://doi.org/10.5194/ESSD-14-4811-2022>.
- [2] IEA. CO2 Emissions in 2022. Paris: 2023.
- [3] Aktar AK, Taşcıkaraoğlu A, Gürleyük SS, Catalão JPS. A framework for dispatching of an electric vehicle fleet using vehicle-to-grid technology. *Sustainable Energy, Grids and Networks* 2023;33:100991. <https://doi.org/10.1016/J.SEGAN.2022.100991>.
- [4] Wang K, Xie Y, Zhang W, Cai H, Liang F, Li Y. Research on optimal dispatch of distributed energy considering new energy consumption. *Energy Reports* 2023;10:1888–98. <https://doi.org/10.1016/J.EGYR.2023.08.040>.
- [5] Hytowitz RB, Hedman KW. Managing solar uncertainty in microgrid systems with stochastic unit commitment. *Electric Power Systems Research* 2015;119:111–8. <https://doi.org/10.1016/J.EPSR.2014.08.020>.
- [6] Iqbal S, Sarfraz M, Ayyub M, Tariq M, Chakraborty RK, Ryan MJ, et al. A Comprehensive Review on Residential Demand Side Management Strategies in Smart Grid Environment. *Sustainability* 2021, Vol 13, Page 7170 2021;13:7170. <https://doi.org/10.3390/SU13137170>.

- [7] Mohseni S, Brent AC, Burmester D. A demand response-centred approach to the long-term equipment capacity planning of grid-independent micro-grids optimized by the moth-flame optimization algorithm. *Energy Convers Manag* 2019;200:112105. <https://doi.org/10.1016/J.ENCONMAN.2019.112105>.
- [8] Hosseini SE, Ahmarinejad A, Tabrizian M, Bidgoli MA. Resilience enhancement of integrated electricity-gas-heating networks through automatic switching in the presence of energy storage systems. *J Energy Storage* 2022;47:103662. <https://doi.org/10.1016/J.EST.2021.103662>.
- [9] Liu S, Gao Y, Yang H, Xie X, Ma Y. Optimal guidance strategy for flexible load based on hybrid direct load control and time of use. *Global Energy Interconnection* 2023;6:297–307. <https://doi.org/10.1016/J.GLOEI.2023.06.004>.
- [10] Choi SY, Lee KY. Direct Load Control System Design Using Active Database. *IFAC Proceedings Volumes* 2014;47:9936–41. <https://doi.org/10.3182/20140824-6-ZA-1003.02489>.
- [11] Chen Y, Yin R. Estimating electricity saving-potential in small offices using adaptive thermal comfort 2022.
- [12] Chen Y-T, Setiawan RJ. Energy Saving Solution for Welding Process: A SME Case Study of Fume Extractor 2023:720–5. <https://doi.org/10.1109/ICKII58656.2023.10332684>.
- [13] Sharma A, Jupalle H, Naddikatu RR, Velu A, Whig P. AI Application for the Sustainable Development to Reduce Carbon Footprint. 2021 5th International Conference on Information Systems and Computer Networks, ISCON 2021 2021. <https://doi.org/10.1109/ISCON52037.2021.9702511>.
- [14] van Putten BJC, Mahdavi N, Braslavsky JH. An Analytical Model for Demand Response of Variable-Speed Air Conditioners. *IFAC-PapersOnLine* 2018;51:426–31. <https://doi.org/10.1016/J.IFACOL.2018.11.740>.
- [15] Erenoğlu AK, Şengör İ, Erdiñç O, Taşcıkaraoğlu A, Catalão JPS. Optimal energy management system for microgrids considering energy storage, demand response and renewable power generation. *International Journal of Electrical*

Power & Energy Systems 2022;136:107714.
<https://doi.org/10.1016/J.IJEPES.2021.107714>.

[16] Yilmaz S, Cuony P, Chanez C. Prioritize your heat pump or electric vehicle? Analysing design preferences for Direct Load Control programmes in Swiss households. *Energy Res Soc Sci* 2021;82:102319.
<https://doi.org/10.1016/J.ERSS.2021.102319>.

[17] Yilmaz S, Chanez C, Cuony P, Patel MK. Analysing utility-based direct load control programmes for heat pumps and electric vehicles considering customer segmentation. *Energy Policy* 2022;164:112900.
<https://doi.org/10.1016/J.ENPOL.2022.112900>.

[18] Moradi-Sarvestani S, Jooshaki M, Fotuhi-Firuzabad M, Lehtonen M. Incorporating direct load control demand response into active distribution system planning. *Appl Energy* 2023;339:120897.
<https://doi.org/10.1016/J.APENERGY.2023.120897>.

[19] Rama Curiel JA, Thakur J. A novel approach for Direct Load Control of residential air conditioners for Demand Side Management in developing regions. *Energy* 2022;258:124763. <https://doi.org/10.1016/J.ENERGY.2022.124763>.

[20] Oprea SV, Bâra A. Edge and fog computing using IoT for direct load optimization and control with flexibility services for citizen energy communities. *Knowl Based Syst* 2021;228:107293.
<https://doi.org/10.1016/J.KNOSYS.2021.107293>.

[21] Tang R, Wang S, Yan C. A direct load control strategy of centralized air-conditioning systems for building fast demand response to urgent requests of smart grids. *Autom Constr* 2018;87:74–83.
<https://doi.org/10.1016/J.AUTCON.2017.12.012>.

[22] Wang S, Tang R. Supply-based feedback control strategy of air-conditioning systems for direct load control of buildings responding to urgent requests of smart grids. *Appl Energy* 2017;201:419–32.
<https://doi.org/10.1016/J.APENERGY.2016.10.067>.

- [23] Dai M, Li H, Wang S. Multi-agent based distributed cooperative control of air-conditioning systems for building fast demand response. *Journal of Building Engineering* 2023;77:107463. <https://doi.org/10.1016/J.JOBE.2023.107463>.
- [24] Adhikari R, Pipattanasomporn M, Rahman S. An algorithm for optimal management of aggregated HVAC power demand using smart thermostats. *Appl Energy* 2018;217:166–77. <https://doi.org/10.1016/J.APENERGY.2018.02.085>.
- [25] Neglia G, Giarré L, Tinnirello I, Di Bella G. Teletraffic engineering for direct load control in smart grids. *Sustainable Energy, Grids and Networks* 2018;16:167–76. <https://doi.org/10.1016/J.SEGAN.2018.07.007>.
- [26] Zhang F, de Dear R. Application of Taguchi method in optimising thermal comfort and cognitive performance during direct load control events. *Build Environ* 2017;111:160–8. <https://doi.org/10.1016/J.BUILDENV.2016.11.012>.
- [27] Hlalele TG, Zhang J, Naidoo RM, Bansal RC. Multi-objective economic dispatch with residential demand response programme under renewable obligation. *Energy* 2021;218:119473. <https://doi.org/10.1016/J.ENERGY.2020.119473>.
- [28] Prathapaneni DR, Detroja K. Optimal design of energy sources and reverse osmosis desalination plant with demand side management for cost-effective freshwater production. *Desalination* 2020;496:114741. <https://doi.org/10.1016/J.DESAL.2020.114741>.
- [29] Lankeshwara G, Sharma R, Yan R, Saha TK. Control algorithms to mitigate the effect of uncertainties in residential demand management. *Appl Energy* 2022;306:117971. <https://doi.org/10.1016/J.APENERGY.2021.117971>.
- [30] Jabian ME, Funaki R, Murata J. Load Shedding Optimization Considering Consumer Appliance Prioritization Using Genetic Algorithm for Real-time Application. *IFAC-PapersOnLine* 2018;51:486–91. <https://doi.org/10.1016/J.IFACOL.2018.11.750>.
- [31] Oprea SV, Bâra A. Edge and fog computing using IoT for direct load optimization and control with flexibility services for citizen energy communities.

Knowl Based Syst 2021;228:107293.
<https://doi.org/10.1016/J.KNOSYS.2021.107293>.

[32] Phangbertha LN, Fitri A, Purnamasari I, Muliono Y. Smart Socket for Electricity Control in Home Environment. *Procedia Comput Sci* 2019;157:465–72. <https://doi.org/10.1016/J.PROCS.2019.09.002>.

[33] Li M, Gu W, Chen W, He Y, Wu Y, Zhang Y. Smart Home: Architecture, Technologies and Systems. *Procedia Comput Sci* 2018;131:393–400. <https://doi.org/10.1016/J.PROCS.2018.04.219>.

[34] Ahmed MS, Mohamed A, Khatib T, Shareef H, Homod RZ, Ali JA. Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. *Energy Build* 2017;138:215–27. <https://doi.org/10.1016/J.ENBUILD.2016.12.052>.

[35] Pawar P, Vittal K P. Design and development of advanced smart energy management system integrated with IoT framework in smart grid environment. *J Energy Storage* 2019;25:100846. <https://doi.org/10.1016/J.EST.2019.100846>.

[36] Zou B, Peng J, Yin R, Luo Z, Song J, Ma T, et al. Energy management of the grid-connected residential photovoltaic-battery system using model predictive control coupled with dynamic programming. *Energy Build* 2023;279:112712. <https://doi.org/10.1016/J.ENBUILD.2022.112712>.

[37] Chiu MC, Kuo TC, Liao HT. Design for sustainable behavior strategies: Impact of persuasive technology on energy usage. *J Clean Prod* 2020;248:119214. <https://doi.org/10.1016/J.JCLEPRO.2019.119214>.

[38] Ocampo Batlle EA, Escobar Palacio JC, Silva Lora EE, Martínez Reyes AM, Melian Moreno M, Morejón MB. A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: Case study, Federal University of Itajubá (UNIFEI). *J Clean Prod* 2020;244:118551. <https://doi.org/10.1016/J.JCLEPRO.2019.118551>.

- [39] He Z, Hong T, Chou SK. A framework for estimating the energy-saving potential of occupant behaviour improvement. *Appl Energy* 2021;287:116591. <https://doi.org/10.1016/J.APENERGY.2021.116591>.
- [40] Department for Energy Security and Net Zero. Greenhouse gas reporting: conversion factors 2023. GOVUK 2023. <https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2023> (accessed July 5, 2023).
- [41] Wang Z, Luther M, Horan P, Matthews J, Liu C. Technical and economic analyses of PV battery systems considering two different tariff policies. *Solar Energy* 2024;267:112189. <https://doi.org/10.1016/J.SOLENER.2023.112189>.
- [42] Ofgem. Energy price cap. Ofgem 2023. <https://www.ofgem.gov.uk/energy-price-cap> (accessed February 4, 2024).

6 Integrated energy management system for off-grid solar energy application⁵

Abstract:

The ability to properly manage and maximise renewable energy sources is crucial to sustainably scaling their usage and share in the grid. This paper develops a comprehensive integrated energy management system (IEMS) that combines Supply-Demand Side Management to predict and control the use of solar energy, while incorporating the time-of-use (TOU) criteria, direct load control (DLC) and generator control (GC) schemes; the main objective of the IEMS framework is to increase photovoltaic (PV) usage in off-grid applications. Firstly, the three-step solar energy forecasting approach is proposed to predict the next-day PV generation by combining low-level data fusion of solar-dependent weather variables into regression models, selecting the best performing model via the lowest root mean square error, and scheduling the load based on TOU criteria. Finally, the IEMS decision algorithm is developed to adjust for forecast errors, initiate DLC to reduce excess load, and/or increase supply by calling up the diesel generator (DG). The proposed strategy is tested by performing a techno-economic analysis of the cost and CO₂ emissions over a 20-year period on three system configurations: DG, DG/PV, and DG/PV/IEMS; a case study of an off-grid residential load supplied by a 7.5 kVA DG and 10 kW PV supply is considered. The results show that the proposed method reduced the lifetime expenditure costs and CO₂ emissions by 44% and 46% respectively when compared to the DG configuration, and by 8% and 9% in the DG/PV configuration. More importantly, it increases the PV usage of the DG/PV by over 113%. A further analysis of the IEMS is conducted by including the effects of customer behaviour (CB) using a 14% override rate, with 80% preconditioning and 20% rebounding.

⁵ This chapter is a paper authored by Tolulope Falope, Liyun Lao, Da Huo, & Boyu Kuang published in the journal of Sustainable Energy, Grids and Networks, 2024, <https://doi.org/10.1016/J.SEGAN.2024.101449>.

A fully automated DG/PV/IEMS with 100% customer compliance outperformed DG/PV/IEMS with customer control, proving that overrides reduce the value of the DLC program. The versatility of the IEMS framework suggests that it can be used with different supply-demand responses, as well as adapted for on-grid applications.

Keywords:

Integrated energy management system, Solar energy forecasting, Supply side management, Demand side management, Time-of-use, Direct load control,

6.1 Introduction

With a yearly increase in global energy consumption and consequently greenhouse gas (GHG) emissions, countries are continuously turning to renewable energy sources (RES) because they are abundant and accessible with little negative environmental impact [1]. Building renewable-based generating plants is key to solving carbon emissions. However, it is not the only solution. In fact, more emphasis should be given to efficiently managing the energy from RES rather than building additional generating plants. A further complexity to the problem of energy management in the system is that RES like wind and solar, are very weather dependent which leads to intermittent energy supply [2] and grid imbalance. To solve this problem, an energy management system (EMS) is needed.

One of the functions of an EMS is to maintain the supply-demand balance by optimally scheduling power and/or load sources to achieve grid stability [3]. The EMS uses a system of methods or strategies to modify, monitor, control, improve and maximise the usage of energy, which will significantly reduce operating expenses [4]. According to research studies, the EMS is divided into predictive energy management system (PEMS) and real-time energy management systems (REMS) [5]. PEMS predicts electricity usage or supply

using previous historical data to determine optimally, how to manage load demand or dispatch energy respectively. REMS employs real-time algorithms to delay or reduce the use of controllable appliances during low supply or high demand periods. If the PV generation is historic (half-hourly, hourly, or longer) and is not in real-time, REMS can be replaced with corrective energy management systems (CEMS). According to Falope et al [6], today's complex grid necessitates an advancement of the EMS into an integrated energy management system (IEMS). IEMS is a system that combines PEMS and REMS/CEMS controls by initiating supply and demand responses to balance the grid's power supply and load and uses technological and communication advancements to maintain multi-source or multi-energy systems. In other words, the IEMS effectively combines three components: the ability to forecast load or supply, control of load demand or demand side management (DSM), and control of power supply or supply side management (SSM).

The physical methods of PV forecasting contain three different methods which are satellite-imaging models, sky images, and numerical weather predictions (NWP) [7]. NWP is one of the most prominent methods used for day-ahead forecasts [8,9]. The term "supply side management" (SSM) describes the actions done to ensure that energy is generated, converted, transmitted, and distributed efficiently. [10]. SSM techniques include battery energy storage system (BESS) integration [11], optimizing and synchronizing alternative or back-up power [12], dispatch algorithm for multiple energy supply sources [13] and so on. A Diesel generator (DG), automatically or manually controlled, may be used to meet peak loads in cases of BESS depletion, rapid changes in load, or RES intermittency [14]. Demand Side Management (DSM) and Demand Response (DR) are sometimes used interchangeably to describe programs that influence a customer's electricity usage. DSM refers to methods that modify an electrical grid's load profile in a way that benefits both demand and supply [15]. It focuses on shifting elastic loads to times when the utility charges less. DR is a subset of DSM and is broadly classified as methods that encourage customers to reduce their short-term energy use in response to a price signal [11]. Some examples of DR are time-of-use (TOU), real-time pricing (RTP) and direct load control (DLC).

The time-of-use (TOU) tariff, also known as time-varying tariffs, is a pricing structure that is time-dependent and charges per kWh of energy [16]. Depending on energy use—during peak, average, and off-peak hours—the cost of power fluctuates throughout the day. In another context, TOU refers to a customer's preferred time of load dispatch based on parameters like cost, type of energy supplied and can be used to either set the tariff, determine peak and off/peak load schedules, or perform a combination of both. Using smart metres and switches, DLC is a response in which a utility provides financial incentives to customers to reduce their electricity consumption [17]. DLC implementation is appropriate for loads that are flexible and controllable and have a significant impact on consumption. Customers have a choice to either accept or override DLC signals, with some DLC programs charging a penalty for non-compliance [18]. Examples of DLC loads are washing machines, dish washers, air conditioners and water heaters.

Customer behaviour (CB) and response to DR programs is crucial in understanding its success. A customer can override a DR command or choose to accept. Customers who choose to override tend to lose whatever financial incentive is offered for curtailing consumption. With 100% customer compliance, we classify it as a fully automated system. Two events created by DR programs like DLC and TOU are called preconditioning (before DLC event) and rebounding (after DLC event). Preconditioning and rebounding are customer behaviours that result in an increase in the customer's load consumption before and/or after a DLC event or during a TOU peak period to compensate for load curtailment, higher usage costs or threat of penalties [19].

IEMS are compatible with different power system configurations and can exist with or without storage, supplying both on-grid or off-grid applications. The choice between BESS and an alternate energy source like DG depends on cost, reliability, and environmental impact. BESS tend to have higher up-front costs than other hybrid alternatives for large installations. As time passes, they also lose the capacity to keep a charge. Due to this unreliability and high cost, BESS is the weakest component of the PV system [20–22]. In an EMS study by

Srikranjana et al. [23], solar PV integrated with an EMS had the highest benefits when considering net present value, internal rate of return, and pay back compared to solar alone, and solar/EMS/BESS. This occurred because the cost of the battery outweighed the combined savings from exported PV energy, savings from efficient energy usage via the EMS, and savings from self-consumption. In most cases, the low compensation rate of PV exports in most countries, combined with a lower EMS implementation cost versus BESS cost, is a major contributing factor to using a system without BESS. Additionally, the ability to recycle spent batteries, especially in developing countries, is non-existent. High investment and replacement costs, chemical pollution, short-life spans have led researchers to explore hybrid configurations without storage. There is also a case for looking at IEMS performance in off-grid applications. According to a 2021 report by the International Renewable Energy Agency (IRENA), the capacity of off-mini grid installations was over 450,000 MW, which is separate from other solar applications like solar home systems, solar pumps, and solar lights [24]. There are over 900 million people in the world without electricity access [25] who are expected to be served by off-grid energy applications. Poorly managed energy systems will lead to energy waste and high GHG emissions. Off-grid applications are also more expensive than on-grid due to storage requirements and can sometimes be a deterrent to investors because of longer returns on investment (ROI). Therefore, reducing wastage and optimizing PV usage in off-grid applications is crucial to solar energy adoption.

Sustainable and reliable energy, financial returns and energy independence are the biggest considerations when developing an EMS that integrates energy from RES into the conventional grid. In terms of cost, the right combination of forecasting methodology, generation units, storage and tariffs within the EMS will result in the best system design. Using an EMS is one of the ways to achieve solar energy integration and optimization, because of its capacity to balance demand and supply considering the intermittency of solar energy. A key question is how can an EMS integrated with PV achieve global system optimisation in terms of operation cost and GHG emission?

According to research, different EMS architectures have been developed. An EMS proposed by Javed et al. [26] incorporates solar panels, wind turbines, energy storage system (ESS), DG, and the utility to reduce the operational expenses and costs of a university. The EMS strategy focused on increasing usage from the RES while employing the combination of TOU pricing and ESS as demand-based responses to optimise the charge and discharge of the batteries based on favourable tariff periods.

In a study by Ouédraogo et al. [27], a PV/battery integrated system supplying a building and an EV is considered. The EMS framework includes decision rules that account for forecasted PV production, electricity price variations and battery degradation costs. The EMS works on a bi-seasonal strategy, separating the months of the year into winter (November to March) and Summer (April to October). During the Winter, energy supply is prioritised in the order of PV, grid and then battery. In the summer, the priority is based on TOU, with the EMS drawing electricity from the grid when the cost is low and using the battery when the cost is high.

In another study by Cupples et al. [28], the EMS takes as its input meteorological data of global horizontal irradiance and wind speed to predict PV and wind power production. Through a smart meter, the EMS obtains RTP tariffs for minimization in the optimisation process and uses a cost-benefit model to aggregate all assumptions, predictions, and system parameters in an economic sensitivity analysis. Using an office building of 3 square meters, the EMS realized savings of \$465 for the month of January 2021.

Mohandes et al. [29] develops a new EMS that uses multiplicative weights update to improve an ensemble of deep neural network-based PV forecasts, optimally size a hybrid PV/BESS system, and reduce fluctuation of supply to the grid from RES.

George-Williams et al. [30] uses a Monte-Carlo based framework to model and simulate EV charging hubs fitted with PV, bi-directional charging stations (vehicle-to-grid), BESS, and dispatchable hydrogen. The EMS accounted for PV module degradation, power losses, varying loads, price inflation and charge point

diversity in order to optimally size smart hubs, preserve battery life, and maximize revenue while ensuring driver satisfaction. The EMS achieved nearly 100% in energy self-consumption for the site in the case study and produced between 1.7% to 3.1% more revenue than without the EMS. In another vehicle-to-grid system, Khizir et al. [31] proposes an EMS for peak-load management that includes PV, BESS and the grid. The EMS reduces peak load by 50% and energy cost by 27.3%.

In another study, Anusha et al. [32] evaluates an EMS for three scenarios: normal operating conditions, grid outage and a total blackout. The EMS manages a micro-grid that includes PV, wind turbine, BESS and a DG versus the base case that uses only DG. Considering a case study of 20 years and 5% interest rate, the base case will cost a total of ₹644,346,341.25 when considering capital, replacement, fuel, maintenance, and salvage costs. It will also result in a CO₂ emission level of 861,401 kg/year. Alternatively, the EMS optimised for 100 kW PV, 500 kW converter, 500 kW DG, 100 kWh Li-ion, and generic flat plate PV, will result in a net present cost of ₹7,958,975,30.

Ali et al. [33] developed an EMS that uses PV and a combined cooling and heating power plant with the national grid. The EMS uses an improved flower pollination technique combined with mixed-integer linear programming to create a hybrid algorithm. Loads are also separated into curtailable, shiftable, and flexible thermal loads. The execution time, cost, and GHG emissions drop by 35.064%, 8.98% and 10.81% respectively. Conditional value at risk was used to combat load loss during poor weather conditions and increased the energy retaining capability of thermal energy storage and electric vehicle by 53.34% and 28.56% respectively.

In a similar study, Kichou et al. [34] investigates a hybrid system that combines PV, a combined heat and power plant, BESS and the utility grid to power a building. Based on load profiles, a control strategy was developed to limit power drawn from the grid from 140 kW to 120 kW with estimated savings of almost 4000 euros/year.

For a stand-alone PV power system, Ozden et al. [35] proposes an EMS optimised with dynamic knapsack and multi-agent devices, including a source agent to determine state-of-charge and load agents that measure instantaneous power consumption. The EMS uses a fuzzy logic-based decision maker to effectively schedule load based on the available energy supply, with priority given to critical loads.

Rochd et al. [36] proposes an EMS with PV, BESS and utility inputs, that uses a multi-objective optimization algorithm to simultaneously maximize user comfort and minimize energy costs based on user preference, the electricity price, and PV generation forecasting. It achieves this through scheduling and control of flexible loads and generation dispatch. A summary of these EMS studies is listed in **Error! Reference source not found..**

Table 6-1 Comparison of Studies on Various EMS Frameworks

Studies	PV Forecasting	SSM		DSM		Financial Analysis	GHG Emissions	CB
		PV	Generator Dispatch	TOU	DLC			
Srikranjanapert et al. [23]	×	✓	×	✓	×	✓	×	×
Harajli et al. [37]	×	✓	✓	×	×	✓	✓	×
Ali et al. [33]	×	✓	×	✓	✓	×	✓	×
Mahmud et al. [31]	×	✓	×	✓	×	✓	×	×
Kichou et al. [34]	×	✓	×	×	×	✓	×	×
Ozden et al. [35]	×	✓	×	×	✓	×	×	×
Rochd et al. [36]	✓	✓	×	✓	✓	×	×	×
Javed et al. [26]	×	✓	✓	✓	×	✓	✓	×
Ouédraogo et al. [27]	✓	✓	×	✓	×	×	×	×
Anusha et al. [32]	×	✓	✓	×	×	✓	×	×
Cupples et al. [28]	✓	✓	×	×	×	✓	×	×
Mohandes et al. [29]	✓	✓	×	×	×	×	×	×
George-Williams et al. [30]	×	✓	×	×	×	×	×	×
Dinh et al. [38]	×	✓	×	×	✓	×	×	×
Silva et al. [39]	✓	✓	×	✓	✓	×	×	×
Ochoa et al. [40]	×	✓	×	×	×	×	×	×
Sharda et al. [41]	✓	✓	×	✓	×	×	×	×
Tabrizi 2022 [42]	×	✓	✓	✓	×	×	×	×
This Study	✓	✓	✓	✓	✓	✓	✓	✓

As inferred in **Error! Reference source not found.**, there is no study that has investigated the incorporation of SEF, GC, TOU, DLC into an IEMS for off-grid applications. Though a few studies considered the financial and environmental impact of their systems, none of them included the impact of CB, which is crucial for the successful implementation of any EMS. It is clear to see that more comprehensive EMS frameworks that incorporate advanced forecasting models, DR strategies, and CB analysis are needed. Such integration would enable more accurate predictions of energy supply and demand, better management of energy resources, and enhanced system reliability. To address the research gap, this paper proposes the following key contributions:

1. An IEMS that optimally schedules and dispatches PV and DG power supply with the help of a three-step solar energy forecasting model, TOU and DLC responses, prioritizing PV while making sure supply meets demand at every time-step.
2. Investigating CB in response to TOU and DLC signals using customer overrides, preconditioning and rebounding percentages.
3. Techno-economic analysis of the IEMS and its impact on CO₂ emissions to determine which hybrid configuration is better for PV integration.

The framework for the proposed IEMS is shown in **Error! Reference source not found.**

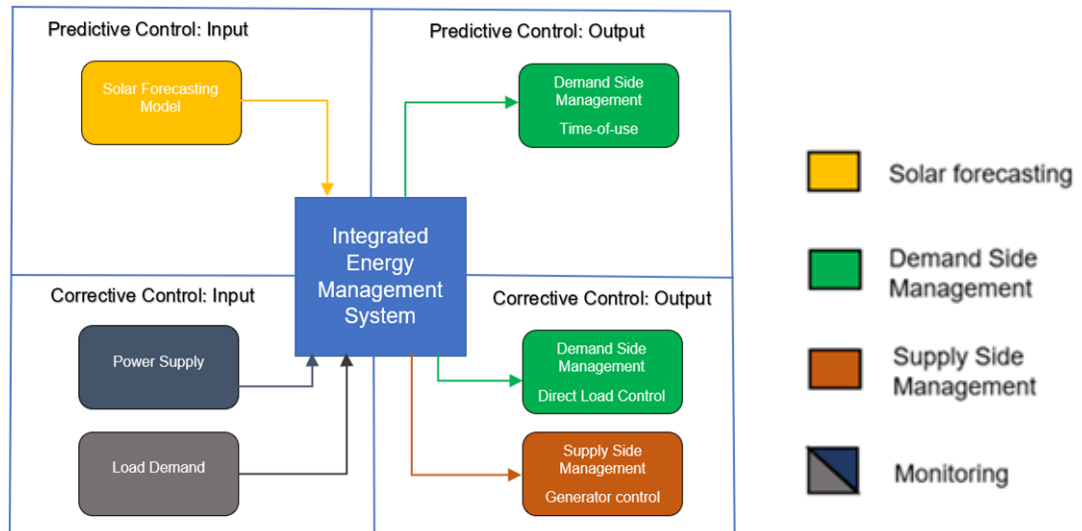


Figure 6-1 Proposed Integrated Energy Management System Framework

The IEMS will be investigated using a home supplied by a combination of a 10 kW PV and a 7.5 kVA DG. One-year residential load profile and historic PV generation were obtained and used for the analysis. Two hybrid off-grid configurations, DG/PV and DG/PV/IEMS will be evaluated against a base-case scenario supplied solely by DG, to determine which configuration is better and to test the performance of the IEMS and its impact on cost and the environment.

This paper is organized as follows. Section 6.2 provides a discussion of methodology and methods used in operating the IEMS. Section 6.3 highlights case study and scenario assumptions, including parameters used in assessing the performance of the IEMS. Section 6.4 discusses the results, while Section 6.5 presents the conclusions and recommendations for future research.

6.2 Methodology

While **Error! Reference source not found.** gives an overall view of the proposed IEMS framework, Figure 6-2 shows how the IEMS interacts with the power supply and the load. The flow of control starts from obtaining information from the generators and the weather stations. The outputs of this are the PV and generator plants generated power in real-time, and the hourly day-ahead PV generation

forecast. The IEMS is then connected to the external power distribution network via smart meters and to the load via smart plugs or sockets. The smart meter enables the TOU tariff or load schedule to be set based on hourly predicted PV generation, while the smart sockets or plugs allow loads to be interrupted or rescheduled based on the dispatch strategy.

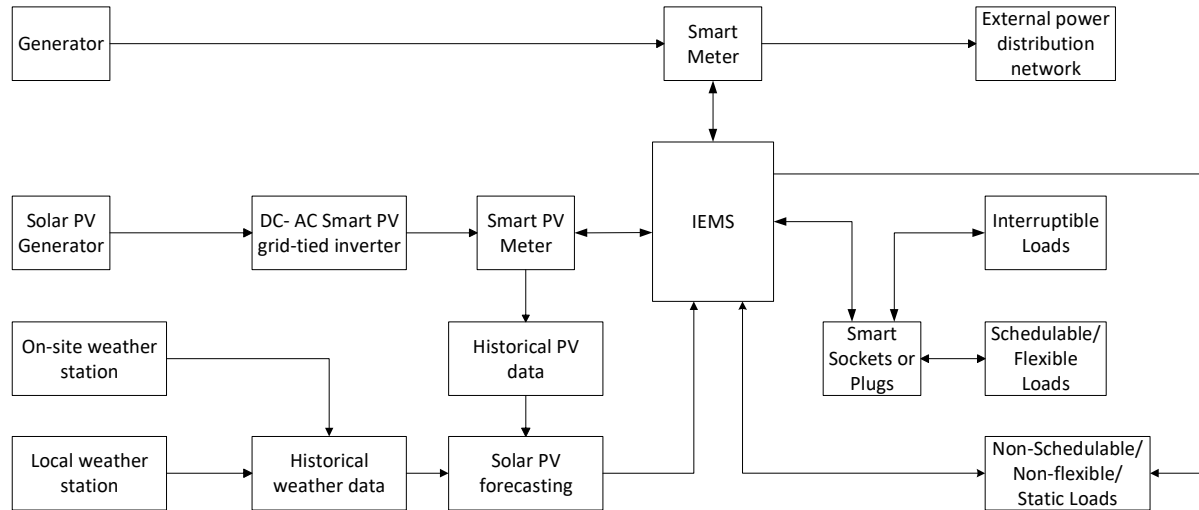


Figure 6-2 Block diagram of proposed IEMS

6.2.1 Three-step forecasting architecture

A three-step forecasting approach for forecasting PV generation is proposed. Using data from a local weather station as well as an on-site one, this method selects weather variables with moderate to strong positive correlation to solar radiation, using them to forecast the production of solar energy. The weather data from the two stations are combined by a technique called low-level data fusion (LLDF), which combines more than one source of raw data to produce new raw data [43]. The data is then normalized and separated into training and testing data samples. The training data sample is used as the input into MATLAB's machine learning toolbox called the Regression Learner App (RLA). The RLA, with a library of over twenty forecasting algorithms, selects the best performing algorithm based on the lowest Root Mean Square Error (RMSE). This approach has been extensively analysed in [44]. A diagram of the proposed three-step forecasting approach is shown in Figure 6-3.

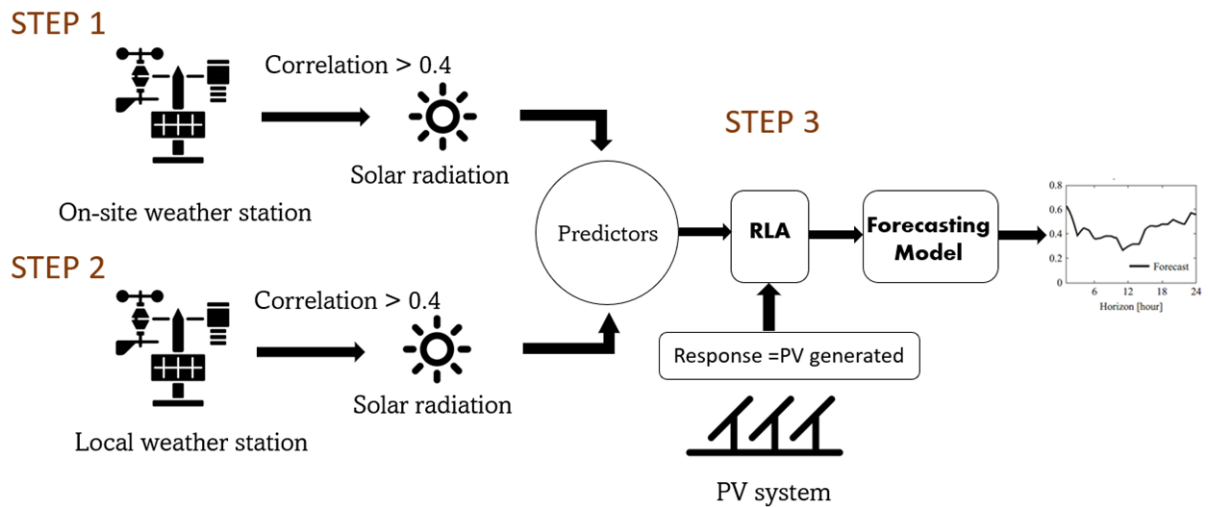


Figure 6-3 Proposed Three-Step Forecasting Approach

In the first step, data recorded in 15 min timesteps by the on-site weather station at Cranfield University comprising of nine meteorological parameters (inclusive of solar radiation) is combined into a correlation matrix with solar radiation as the independent variable due to its importance in PV generation. In the second step, the same solar radiation recorded by the on-site weather station is selected as the independent variable and combined into a correlation matrix with twelve other meteorological parameters recorded by the Met Office Bedford station in hourly timesteps. In the final step, the meteorological parameters from both weather stations with moderate to strong correlations (> 0.4) on the same corresponding hourly timesteps are selected as forecasting model predictors. These weather parameters along with the corresponding hourly PV generated become the inputs into the RLA.

6.2.2 Financial forecasting tools

Discount Factor

The discount factor (DF) is a weighing factor that is multiplied by a future cash flow to express the latter in its present value.

$$DF_n = \frac{1 - (1 + r)^{-n}}{r} \quad (6-1)$$

Where r is the discount factor or interest rate, and n is the time span.

Net Present Value

The total of the present values of all future cash flows is known as the net present value (NPV). NPV is a measure of the value that a project or investment contributes to a developer or business and is used to evaluate a project's profitability [45]. The higher the returns, the higher the NPV.

$$NPV = \sum_{t=0}^n \frac{Cf_n}{1 + r^n} - C_0 = (I_n \times DF_n) - C_n \quad (6-2)$$

Where Cf_n is the future cash flow from period 0 to n , n is the time span, C_0 is the initial cost of investment at time zero, C_n is the capital cost, F_n is the fuel cost, and I_n is the income.

Internal Rate of Return

Internal rate of return (IRR) is used to compare or assess investment profitability [46]. IRR, or the rate at which NPV equals zero, is the discount rate at which the initial investment is equal to the present value of all future cash flows.

$$IRR \sim NPV = \sum_{t=0}^n \frac{Cf_n}{1 + r^n} - C_0 = (I_n \times DF_n) - C_n = 0 \quad (6-3)$$

Life-Cycle Cost Analysis

The life-cycle cost analysis (LCCA) considers the present value of all costs related to an investment over the course of its existence [47]. In other words, the overall cost of ownership over a known duration. This covers the acquisition expenses (capital cost, installation costs), the operations and maintenance costs (O&M), the fuel costs, the replacement costs, the conversion and/or the decommissioning costs at the end of the investment's life [48].

$$LCC = \sum_{t=0}^n \frac{[(C_n + M_n + F_n - R_n) - I_n]}{1 + r^n} = C_n + \{(M_n + F_n) \times DF_n\} - (I_n \times DF_n) \quad (6-4)$$

Where M_n is the operations and maintenance cost, F_n is the fuel cost, R_n is the residual cost and I_n is the income.

Cost-Benefit Analysis

Prudent developers will perform a cost-benefit ratio (CBA) to assess all potential costs and profits that may be generated if the project is finished prior to building a new facility or taking on a new project. Typically, just the recurrent revenue is considered as the benefit when calculating a cost-benefit ratio (CBR).

$$CBR = \sum_{t=0}^n \frac{[(C_n + M_n + F_n - R_n) \div I_n]}{1 + r^n} = \frac{C_n + \{(M_n + F_n) \times DF_n\}}{(I_n \times DF_n)} \quad (6-5)$$

Payback Period

The payback period (PP) is the duration required for an investment to generate sufficient income or cash to cover the initial investment's cost.

$$PP = \frac{\text{Capital costs}}{\text{Annual savings}} \quad (6-6)$$

Levelized Cost of Energy

The levelized cost of energy (LCoE) over the economic lifespan of a system is the proportion of the net present value (NPV) of the total investment (capital costs, fuel expenses, O&M costs) to the net present value (NPV) of the total electricity generated by the system.

$$LCoE = \frac{\sum_{t=1}^n \frac{I_t + M_t + F_t}{(1+r)^t}}{\sum_{t=1}^n \frac{E_t}{(1+r)^t}} = \frac{\{(M_t + F_t) \times DF_n\}}{E_t} \quad (6-7)$$

For the time period t , E_t is the electricity generation, I_t is the investment cost, F_t and M_t are the fuel and O&M costs respectively.

For a better understanding of the forecasting tools and their impact on a project, Table 6-2 has been adapted from definitions summarised in [49].

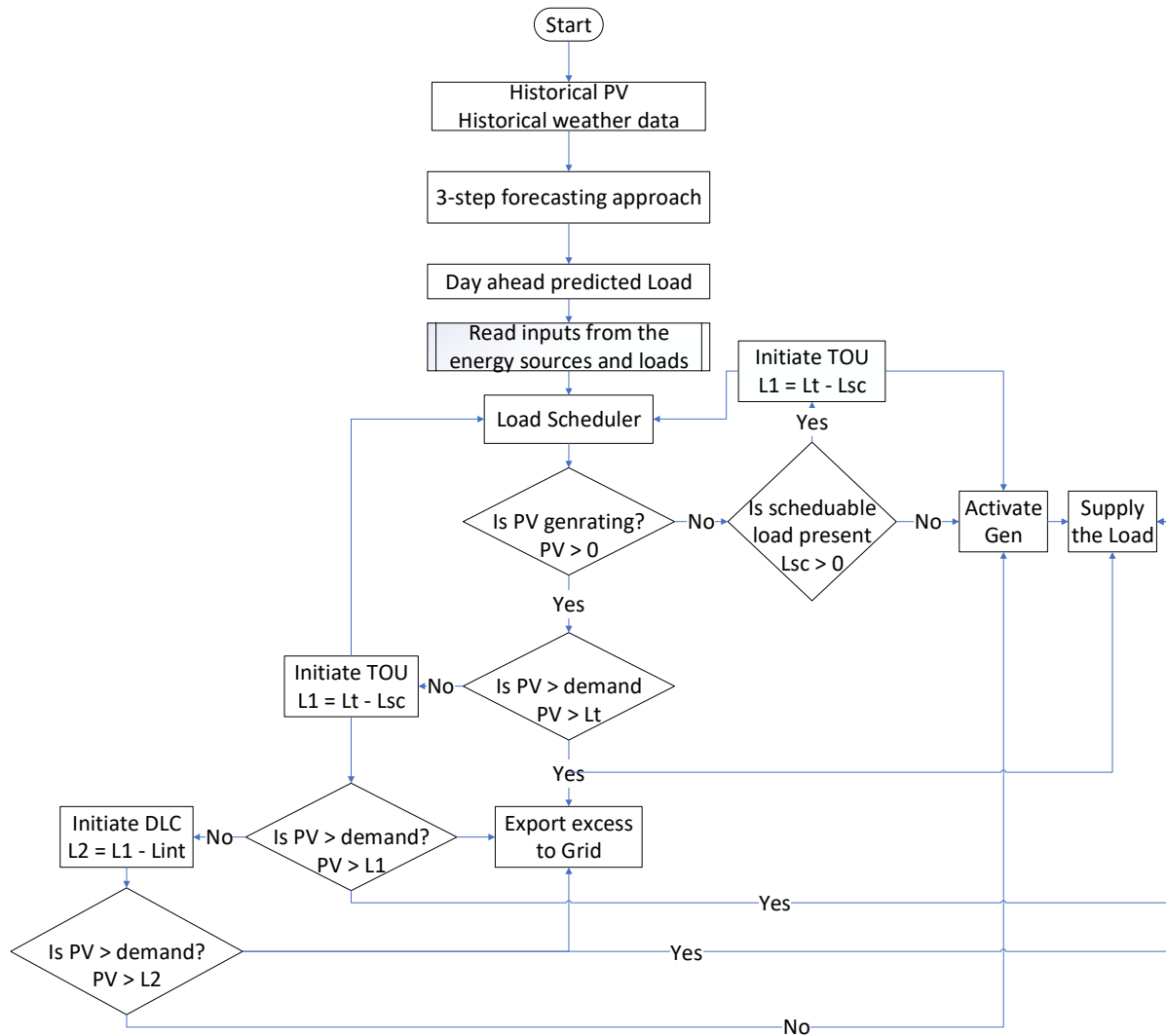
Table 6-2 Financial conditions and definitions

Condition	Definition	Result
$NPV > 0$	Investment adds value	Project may be accepted
$NPV < 0$	Investment subtracts value	Project should be rejected
$NPV = 0$	Investment neither gains nor loses value	Project will break-even. Other criteria needed for decision
$IRR_2 > IRR_1$	Project 2 has a higher IRR than project 1	Project 2 is more desirable
$LCCA > 0$	Fails to generate net income	Project may be rejected
$LCCA < 0$	Generates net income	Project should be accepted
$LCCA = 0$	Breaks even	Adds no monetary value. Other criteria needed for decision
$CBR > 1$	Cost exceeds benefit	Project may be rejected
$CBR < 1$	Benefit exceeds cost	Project should be accepted
$CBR = 1$	Benefit equals cost	Adds no monetary value. Other criteria needed for decision
$PP_2 > PP_1$	Project 2 has a longer PP than project 1	Project 1 could be more desirable
$Cost > LCoE$	Electricity cost exceeds LCoE	Greater return on capital
$Cost < LCoE$	LCoE exceeds electricity cost	Lower return and a possible loss

6.2.3 Dispatch strategies

The IEMS receives as its input a day-ahead solar forecast of the PV system from the proposed three-step forecasting architecture approach described in Section 6.2.1. Based on this forecast, it sets the TOU tariffs/and or load schedule as the first level of control. The IEMS will notify the user through an Internet-of-Things (IoT) enabled environment during load peak periods or periods with low supply. The IEMS with the help of smart meter control, can do a combination of three things: set higher TOU tariffs to reflect periods of low PV supply, notify users about low and peak periods so that they can change/update load priority requests (manually schedule loads), or automatically schedule the load based on PV availability. By doing this, the PEMS function of the IEMS determines the operating schedule of shiftable and interruptible household appliances for the next 24 hours. This is then passed down to the CEMS, which adjusts the schedule due to forecast uncertainty and errors.

For the CEMS, the dual monitoring of customer load and power supply levels serves as its input while GC and DLC are the output as shown in **Error! Reference source not found..** The CEMS regulates the system by checking if the current demand is above the supply level. If this condition is met despite the TOU tariff, the CEMS decisive algorithm will turn-off appliances by using DLC to reduce the load based on customer's demand priority. This is the second level of control. For DLC, the IEMS will use a centralized controller connected to each distributed energy resource in the system via smart plugs or adapters for remote control and access. If this is still not sufficient to curtail the load and balance the supply and demand, then the back-up generator will be called upon to increase the supply. In summary, the IEMS priorities load and energy management before calling up additional supply. Figure 6-4 shows the flowchart of the proposed IEMS.



Where L_t = Total load, L_{sc} = Scheduable load, L_{int} = Interruptible load, L_1 and L_2 = Arbitrary loads

Figure 6-4 Flowchart for proposed IEMS

For off-grid applications without storage, there must be a way to deal with excess or unused PV via dummy loads or by exporting to a local islanded network for consumption by other load sources. In this study, excess PV is exported to a local grid as can be seen in Figure 6-4.

6.2.4 Integrated energy management system rule-based algorithm

6.2.4.1 Load dispatch

As seen in the flowchart in Figure 6-4, the IEMS rule-based algorithm prioritises and maximises the use of PV before activating the generator. It initiates DSM before SSM. Once PV generation has been detected, the IEMS will compare the

load and the PV. If the PV supply is less than the load, the IEMS will reschedule any flexible load. If the load is still greater, the IEMS will initiate DLC and deactivate any interruptible load. If this is still insufficient to reduce the load, then the generator is called as a back-up. The base load is the first to be powered, and along with the schedulable load, must always be powered during any 24-hour period. The interruptible load on the other hand does not always need to be powered. The power to this load is supplied by the DG when there is no PV and will only be powered by the PV if it is sufficient to power the daily total load.

6.2.4.2 Forecast errors

Apart from generally scheduling the load and supply, the IEMS must account for the errors in the PV forecast prediction by adjusting the load schedule to suitable times of actual PV supply. To achieve this, the IEMS runs a high to low priority-based load schedule. It queues schedulable load from the highest kW to the least and assigns the highest schedulable load to the period of the highest predicted PV supply. It assigns the second highest schedulable load to the second highest hour of predicted PV supply and so forth. This flowchart can be seen in Figure 6-5. The IEMS will concurrently look for the earliest time of excess PV and will initiate a schedulable load once the PV supply is sufficient to power both the base load and the schedulable load. In other words, the schedulable load may be powered at any earlier time before its scheduled time if there is enough PV supply to satisfy both the schedulable load and the base load. If there are no periods of excess PV prior to its scheduled time, then the schedulable load will be activated at its original time which is based on the predicted PV. The IEMS looks at the actual PV supply, checks to see if any of the queued loads can be deployed prior to its scheduled time, and will ultimately go back to the default schedule if the actual PV does not exceed the predicted PV. In addition, according to the block diagram in Figure 6-2, the system is always collecting historical PV and weather data to improve the forecasting. The forecasting model takes the actual PV, adds it to its database so the model can be re-trained with new data.

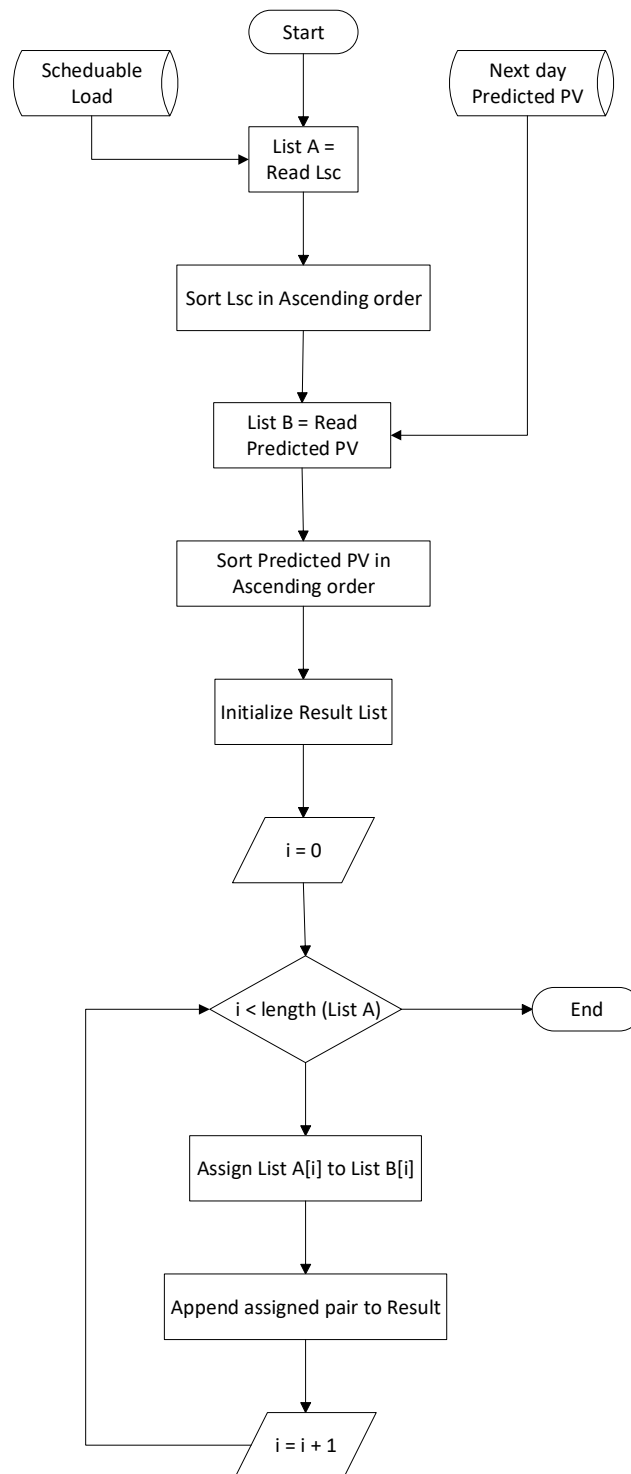


Figure 6-5 Flowchart for assigning schedulable loads

6.3 Case study

6.3.1 Data description

To generate the next day PV forecast, one-year historical weather data from both an onsite weather data station located in Cranfield and a local weather station located in Bedford, as well as one-year historical PV generation for a 10-kW PV system in Cranfield are obtained. One-year UK-based residential load consumption is obtained using an energy data logger. An energy or power data logger is an electronic device that records/logs or monitors electricity consumption, power quality and other power parameters over time. It is crucial in developing the load profile of a customer. In assessing the technical and economic feasibility of installing a hybrid green system for the school of engineering, University of Nairobi in Kenya, Eze et al. [50] uses a power and energy data logger to determine the building's load profile by monitoring the electrical power demand and consumption. Gerber et al. [51] uses a data logger to aggregate and store voltage and current data in a measurement-informed modelling to quantify and compare energy efficiency and power quality of buildings. The aim of the data logger in this case would be to identify and classify different loads and their time of operations. A data logger was used to obtain the energy consumption for the residential house. An example of the one-day load profile is shown in Figure 6-6.

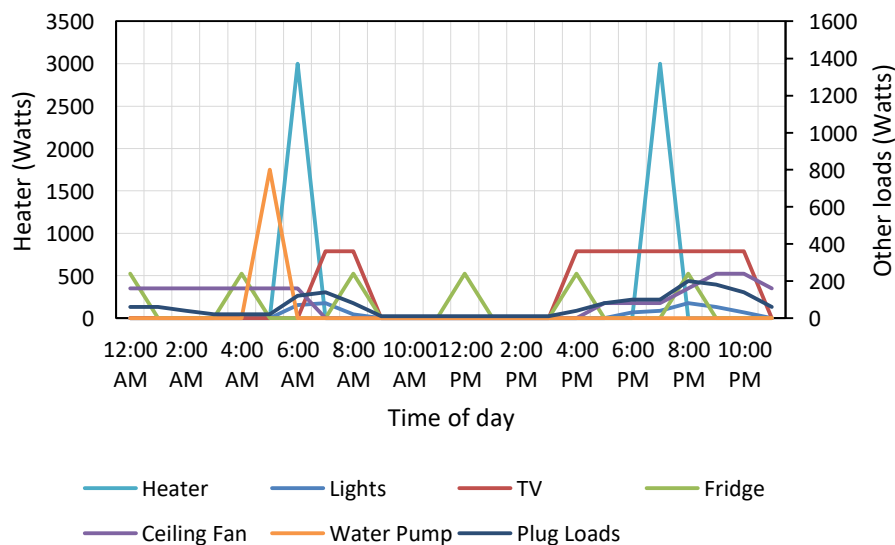


Figure 6-6 One-day load profile of residential house

The daily load is categorized into base/non-schedulable load, schedulable/shiftable/TOU load, and interruptible/Curtailable/DLC load. Interruptible loads can be briefly switched off without causing the user too much discomfort [52]. Schedulable loads are those that can be moved to a more convenient time of operation based on TOU or because of user preference. They are also referred to as flexible or controllable [53]. Non-schedulable or base loads are the highest priority for the customer and cannot be scheduled, interrupted, or delayed. For this study, the plugged loads are the interruptible loads, the heater and water pump are schedulable, and all others make up the base load. Table 6-3 gives the load summary along with the average operating times.

Table 6-3 Load demand summary of residential house

Load Type	Load	Ratings (W)	Quantity	Average operating time (h/day)	Total demand (Wh/day)
Base	Lights	10	8	6	480
	TV	360	1	8	2880
	Fridge	240	1	6	1440
	Ceiling fan	80	3	9	2160
Schedulable	Heater	3000	1	2	6000
	Water Pump	800	1	1	800
Interruptible	Plug Loads	200	Lump	9	1800
Total		4690			15,560

6.3.2 Financial and economic analysis

For these assumptions, costs will be reflected in 2023 prices based on inflation rates and the federal consumer price index (CPI) supplied by the International Monetary Fund (IMF) [54].

6.3.2.1 PV system

In this study, an installed 10 kW PV system acts as a back-up to a DG supplying a monthly average residential load of over 500 kWh. With the aid of a supply meter, daily half-hourly meter readings were taken for the entire year in 2019,

and the average monthly PV output was determined to be over 432 kWh/month. The analysis will consider 20 years based on the life cycle of PV panels [47,48]. Two key terms that must be considered are the discount rate and discount factor. The discount rate determines the value of future cash flow, and it is affected by a country's present economic state [55]. According to literature, discount rate was found to be between 5% and 7.25% [56,57]. In this case study, the unlevered 6% rate will be considered. According to IRENA, the average installed cost for PV worldwide in 2021 was \$867/kW, or the equivalent of £8,000 for a 10 kW PV system adjusted by CPI in 2023. A survey of several vendors revealed a range of between £11,000 to £14,000 [58–60] for the 10 kW PV system. For the purposes of this study, the average of £12,500 will be considered. According to a report by the Electric Power Research Institute (EPRI), the overall budget for PV O&M costs is between \$10 to \$45/kW-year. The report emphasises that for larger utility scale PV systems with a capacity of 1MW and over, the O&M costs are approximately \$10 to \$25/kW-year and are cheaper than smaller units due to economies of scale. Ramasamy et al. [61] give the current benchmark of residential O&M costs without storage at \$28.97/kW-year. With that in mind, we opt for an amount of \$30/kW-year, adjusted with the CPI to £26/kW-year.

6.3.2.2 Diesel generator

The DG was selected based on the worst-case scenario of simultaneously running all the loads at the same time. From Table 6-3, this amounts to approximately 5 kW, or 6.25 kVA using the common industry standard power factor of 0.8 [62]. As the primary source of power, the generator is sized using about 80% of its capacity and should be capable of powering the entire load without back-up. With this design consideration in mind, a 7.5 kVA generator is selected. In terms of cost, several estimations are made based on the DG's capacity. For a 48-kW DG, Das and Zaman [63] list the capital cost as \$370/kW, replacement cost as \$296/kW and O&M costs as \$0.05/h with a lifetime of 15,000h. In a similar cost and price breakdown, Kumar et al. [64] gives the costs as \$665/kW, \$535/kW, \$0.027/h with a lifetime of 15,000h. Ghenai and Bettayeb [65], give the capital and replacement costs of a 100kW generator as \$300/kW, O&M costs as \$0.01/h, fuel cost as \$0.6/litre with a lifetime of 15,000h. In another

study by Ismail et al. [66], the capacity of the DG considered is 7.5 kW and the associated capital cost, fuel cost and fuel inflation rate are \$550/kW, \$0.6/litre and 5% respectively. The above is closer to our rated generator sizing and therefore the capital cost is estimated as £400/kW and replacement cost is £300/kW. The fuel consumption loading is derived from Greaves power generator specification sheet [67]. The diesel CO₂ conversion factor for an average biofuel blend usually obtained from a filling station forecourt is obtained from the UK government website [68]. On 23/05/2023, the daily cost price for ultra-low sulphur diesel fuel [69] was obtained, while an average of the half-hourly PV export rates [70] was also obtained.

6.3.2.3 DLC/TOU programs

DLC and TOU programs use tariff discounts/credits to incentivize customers to remotely shut-off or reschedule their load to an off-peak period. According to a report by the Rocky Mountain Institute [18], DLC programs are typically applied to residential and commercial customers with less than 100 kW demand. Most frequent DLC loads that are controlled are air conditioners (ACs) but may also include water heaters and pool pumps. DLC programme incentive schemes usually consist of a one-time participation payment in addition to fixed monthly payments that are credited to the customer's utility bill. In a study by Kaczmariski et al. [71], the median willingness-to-accept was \$9.50 per month through the summer, or approximately \$30 per year with a confidence interval of 95%. This is similar to a DLC program like PNM Power Saver in New Mexico that offers \$25 per year. In another study by Xu et al. [72], utility providers in the United States offer \$25 to \$100 per year, \$5 to \$20 per month, or 3 cents to \$1 per kWh saved. In a survey of 1482 residents, the authors proved that a \$30 per year incentive boosted accepted rates into the DLC program. In this case study, a benchmark price of £25 per year will be used for the Plug loads when there are no overrides to the schedule between the periods of 5am to 5pm (DLC period). The penalty for non-compliance in overriding the DLC command is at the discretion of the utility operator. For example, during an emergency event, Xcel Energy in Colorado gives a dollar fine of half the annual credit multiplied by the amount of

the online load that is in excess of the contractual interruptible load [18]. In our case, we choose to half the yearly credit for overrides above 10%.

For TOU credits, utilities may offer discount on the normal electricity tariff based on the type of peak period (high, mid or low) or even offer a flat fee based on time of consumption or the consumption profile [73]. For off-peak TOU costs, the Rocky Mountain report cited earlier gives examples of 5% bill reduction and a \$0.10 per kWh for schedulable loads. In a survey of 6 utility providers in the UK, the difference between the peak and off-peak rates for the “Economy 7” electricity tariff for residential customers was 54%. The significant difference helps to incentivize customers to move their bulk load to off-peak periods. With this in mind, we consider a discount of £0.2 per kWh for any load rescheduled to the off-peak period between 5am and 5pm (same as DLC period). The choice of this time-period coincides with PV generation, encouraging customers to operate the bulk of their load (schedulable load) using PV. The total daily schedulable load is 6.8 kWh, which multiplying by £0.2 per kWh equates to a saving of £496 per year.

The Difference between the DG/PV and the DG/PV/IEMS configurations is the IEMS controlled DLC technology for remotely shutting down and rescheduling the load. For our case study, the prices were provided by a leading industry player and include a supervisor (£1100) connected to a gateway hub (£155) that provides the control to the smart sockets. The 200W plug loads will be spread across 5 smart sockets or 10 outlets (£645). Add an installation cost of £300 and the total cost sums up to £2200.

6.3.2.4 Customer behaviour factors

In a study of 6389 households across 21 U.S states and a Canadian province [74], the data showed a mean thermostat override rate of 12.9%. We assume a similar override rate of 14%. Analysing the use patterns of interruptible and schedulable loads based on our historical profile data, an 80% preconditioning rate and 20% rebounding rate were proposed rather than reverse used by Pal et al. [19]. Figure 6-7 shows what the DLC/TOU program looks like while Table 6-4 has a summary of the assumptions broken down into monthly rates for analysis purposes.

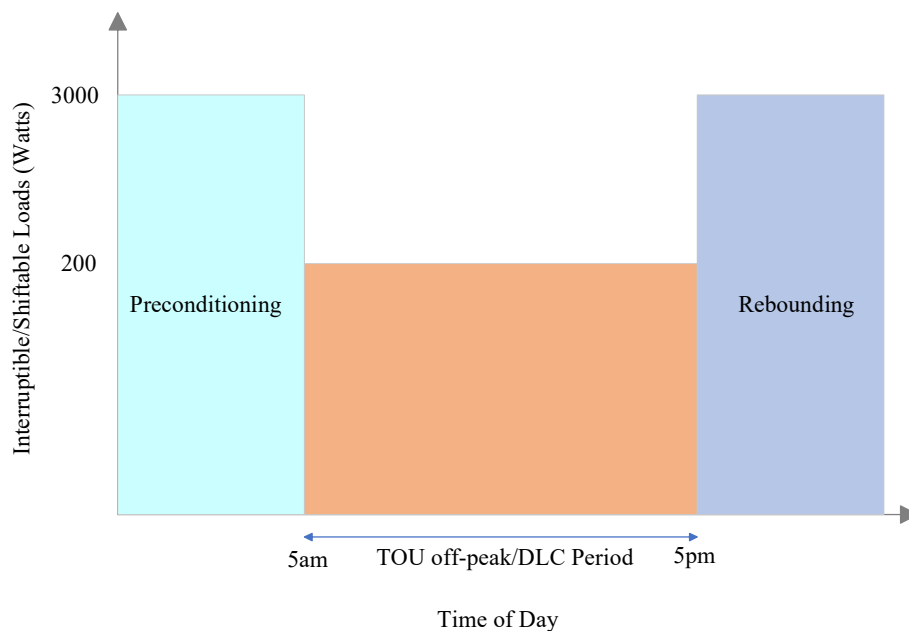


Figure 6-7 Demand load control/Time-of-use program

Table 6-4 Parameter assumptions

Item	PV	DG
Capacity	10 kW	7.5 kVA
Average monthly output	323 kWh/Month	509 kWh/Month
Capital cost	£12,500.00	£400/kW
DLC equipment	£2,200.00	
O & M cost	£2.17/kW-month	£0.14/hour
TOU off-peak discount	£0.2 per kWh	
TOU off-peak/DLC period	5am – 5pm	
DLC credit	£2.08/month	
Load override rate	14%	
Preconditioning rate	80%	
Rebounding rate	20%	
Discount rate	0.5%/month	
Lifespan	240 months	15000 hours
Replacement cost		£300/kW
Fuel cost (diesel)		£1.516/litre
Fuel consumption		100% load = 2.6 (ltr/hr) 75% load = 2.3 (ltr/hr) 50% load = 1.8 (ltr/hr) 25% load = 1.4 (ltr/hr)
Grid export	£63.62/MWh	
CO2 conversion factor	0.0	2.51 kgCO ₂ e/litre

6.4 Results and discussion

To evaluate which of the configurations is better, several variables must be determined. Firstly, the monthly load and PV supply for the residential house for the year 2019 is obtained. In the DG/PV configuration, the yearly PV generation is subtracted from the load to determine the share of DG supply. For the DG/PV/IEMS configuration, the IEMS algorithm highlighted in Section 6.2.4 determines the DG and PV supply outputs. Table 6-5 provides the results.

Table 6-5 Consumption comparison between DG, DG/PV, and DG/PV/IEMS configurations for a residential house in 2019

	DG		DG/PV			DG/PV/IEMS		
Month	DG Load (kWh)	DG hrs	DG Load (kWh)	PV Load (kWh)	DG hrs	DG Load (kWh)	PV Load (kWh)	DG hrs
Jan	734.61	744	674.84	59.78	656	602.99	97.61	629
Feb	590.08	672	513.97	76.11	527	396.08	166.08	506
Mar	487.94	744	434.33	53.61	436	287.34	183.33	418
Apr	446.91	720	372.75	74.17	454	237.87	198.37	467
May	485.81	744	333.74	152.08	434	203.27	275.14	447
Jun	470.14	720	316.15	153.99	399	208.68	268.37	390
Jul	485.81	744	323.49	162.32	403	171.84	312.87	394
Aug	485.81	744	350.24	135.58	434	191.00	290.12	416
Sep	455.85	720	358.82	97.03	450	227.52	219.97	450
Oct	485.81	744	419.83	65.99	434	308.82	162.69	416
Nov	455.85	720	421.45	34.40	420	359.42	79.37	403
Dec	485.81	744	465.28	20.54	434	400.75	63.36	416
Total	6070.44	8760	4984.88	1085.58	5481	3595.59	2317.28	5354
Average	505.87	730	415.41	90.47	457	299.63	193.11	446

6.4.1 Percent share of diesel generator and photovoltaic loads

Table 6-5 shows that DG/PV has a higher combined load generation output (energy supplied from both DG and PV) than the DG/PV/EMS configuration. This is because the IEMS shuts down some interruptible loads without rescheduling them. In addition, the IEMS reduces the yearly generator output by about 28%,

leading to reduced CO₂ emissions and fuel costs. Even more importantly, the IEMS improves the yearly PV usage of the DG/PV configuration by more than 113%, thus achieving the overall aim of integrating more PV into the grid system. Figure 6-8 gives some further insight into the impact of the IEMS on the system.

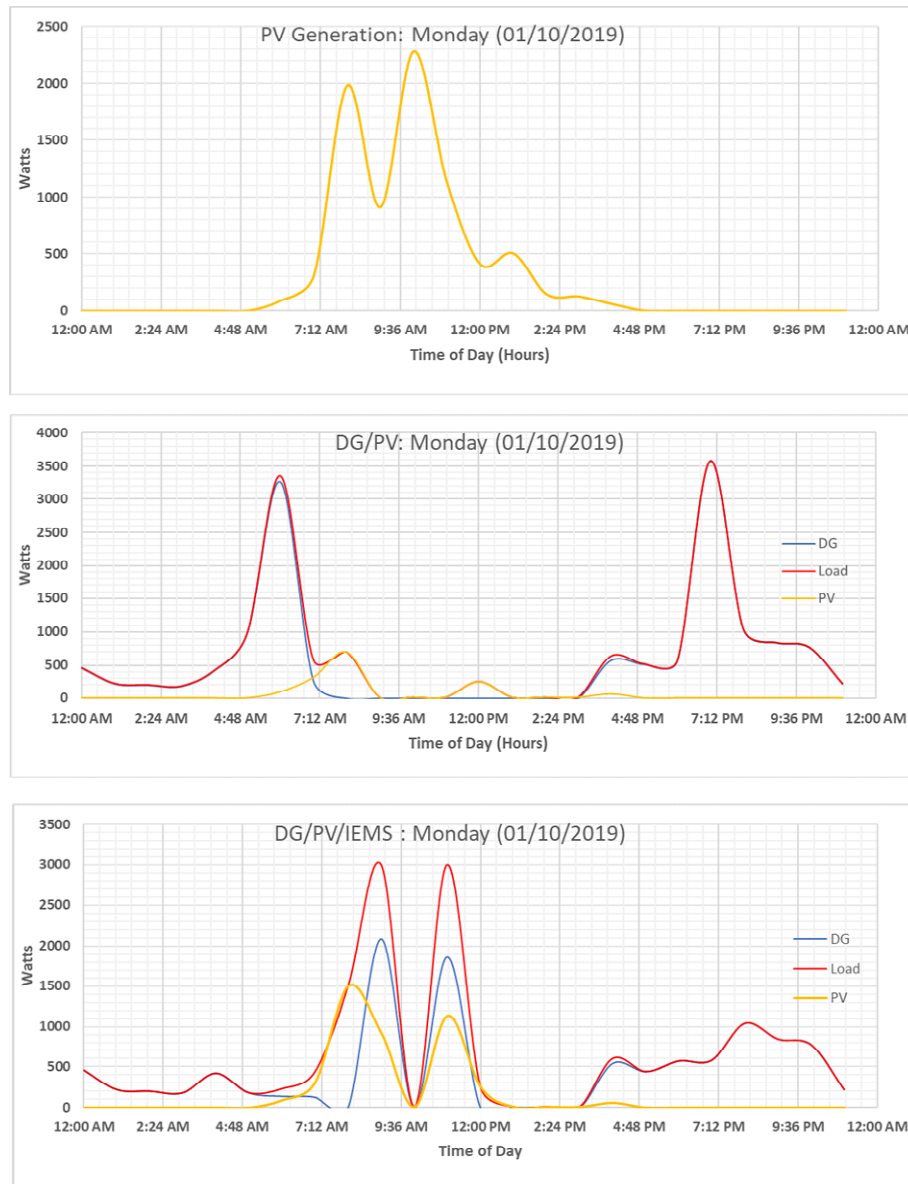


Figure 6-8 Comparison of DG/PV and DG/PV/IEMS configurations servicing the load

The first graph of Figure 6-8 displays the PV generation for the selected day with peak periods between 7am and 12pm. In the second graph showing the DG/PV configuration, the DG follows the load closely and is mostly supplemented by PV. Even though there is excess available PV generation between 7am and 12pm, the major load consumption occurs outside this time frame, between the hours of 4am - 7am and 5pm – 9pm. This means that the excess PV must be exported. In the final graph, the IEMS shifts the load peaks to better coincide with the maximum PV generation. As a result, there is an increase in load usage as can be seen by the bigger load peaks during 7am and 12pm. In addition, these shifted loads are increasingly being powered by PV.

6.4.2 Forecast accuracy and algorithm error correction

Despite slight differences between the hourly forecast and the actual PV generated, the 3-step forecasting algorithm was able to accurately predict when the highest PV generation occurred and correctly assign the heater and water pump to those times. This means that the loads with the highest consumption were powered by PV. In the future, network operators can incentivize customers to schedule their load this way so that the bulk load is shifted to periods of the highest PV generation, thereby reducing CO₂ emissions. Lastly, the forecast error correction algorithm deployed the schedulable load earlier 40% of the time, with the 800W water pump making up the majority. This significant number of times proves that the error correction is effective and useful. Scheduling loads early only occurs when the PV Plant can power the shiftable load more than once a day. In our case, this occurs when the PV plant can generate more than 3000W or 800W multiple hours in a day.

6.4.3 Lifetime expenditure and CO₂ emissions

To determine the fuel costs, the data logger is used to record the hourly load consumption and conversely, the operating hours of the DG can be estimated. The DG loading percentages from Table 6-4 are applied based on the operating hours, and the amount of fuel used each hour can then be determined. The PV export is simply the share of the PV supply not consumed by the load. As this

system does not have any storage, the PV is exported to a local islanded grid. Table 6-6 provides the results.

Table 6-6 Fuel comparison between DG, DG/PV, and DG/PV/IEMS configurations

	Gen	DG/PV		DG/PV/IEMS	
Month	Fuel (L)	Fuel (L)	PV Export (kWh)	Fuel (L)	PV Export (kWh)
Jan	1153.20	689.51	39.13	612.07	1.30
Feb	1014.80	606.76	513.97	538.62	56.25
Mar	1101.39	658.53	277.54	584.57	147.57
Apr	1064.57	690.43	344.75	662.57	219.57
May	1101.39	651.00	498.73	626.20	374.40
Jun	1065.86	602.14	470.56	554.57	353.79
Jul	1101.39	608.04	633.94	551.80	480.90
Aug	1101.39	658.53	514.54	584.57	358.78
Sep	1065.86	683.57	463.17	639.43	339.83
Oct	1101.39	658.53	257.29	584.57	160.59
Nov	1065.86	637.29	153.53	565.71	108.56
Dec	1101.39	658.53	98.28	584.57	55.46
Total	13038.46	7802.85	4265.44	7089.26	2657.00
Average	1086.54	650.24	355.45	590.77	221.42

Again, Table 6-6 shows the yearly PV export of the DG/PV/IEMS configuration is about 38% less than the DG/PV configuration mainly because the IEMS prioritises using PV before exporting it. To further evaluate the performance of the IEMS, both configurations are analysed against the base-case scenario. The base-case scenario involves using only a DG to supply the load. The DG is assumed to supply the load 24 hrs a day, throughout the whole year. Since it will be detrimental for the DG to supply uninterrupted power for every hour of the year, two set of DGs are considered; one for night loads and one for day loads. By using the assumptions of Table 6-4, the capital cost is calculated as £6000. While this is far below the capital costs of the two other configurations, the DG is significantly higher in replacement costs, fuel costs and conversely overall lifetime expenditure. It also gives significantly higher CO₂ emissions over the system lifespan. The Scope 1 emissions, which are direct emissions generated by sources owned by the user, is calculated by multiplying the CO₂ conversion

factor by the total litres of fuel in the year, and the system lifespan which is 20 years.

The replacement cost is calculated based on how many times the DG needs to be replaced in 20 years based on a 15,000 hours-lifespan. To calculate this cost, the total lifespan of the DG, the replacement cost rate from Table 6-4 and the total yearly output from Table 6-5 are needed. Table 6-7 provides the results.

Table 6-7 Cost, Output and CO₂ emissions comparison between DG, DG/PV, and DG/PV/IEMS configurations

Costs & CO₂ (System lifetime)	DG	DG/PV	DG/PV/IEMS
Capital costs	£6,000.00	£15,500.00	£17,700.00
O&M costs	£24,528.00	£18,376.67	£18,269.19
Replacement costs	£26,280	£16,444.04	£16,061
Fuel costs	£395,326.02	£236,582.49	£214,946.33
CO ₂ Emissions	654,531 kg CO ₂ e	391,703 kg CO ₂ e	355,881 kg CO ₂ e

The fuel costs, replacement and O&M costs sum up to the lifetime expenditure cost of each configuration. However, the PV component of the O&M cost is determined by multiplying the O&M cost rate with the PV capacity and the DF. The DF is calculated by using Equation (6-1) and determined to be 139.581. It is important to note that the DF considers the entire lifetime of the system and scales it to a monthly multiplier. Figure 6-9 shows a visual representation of the costs.

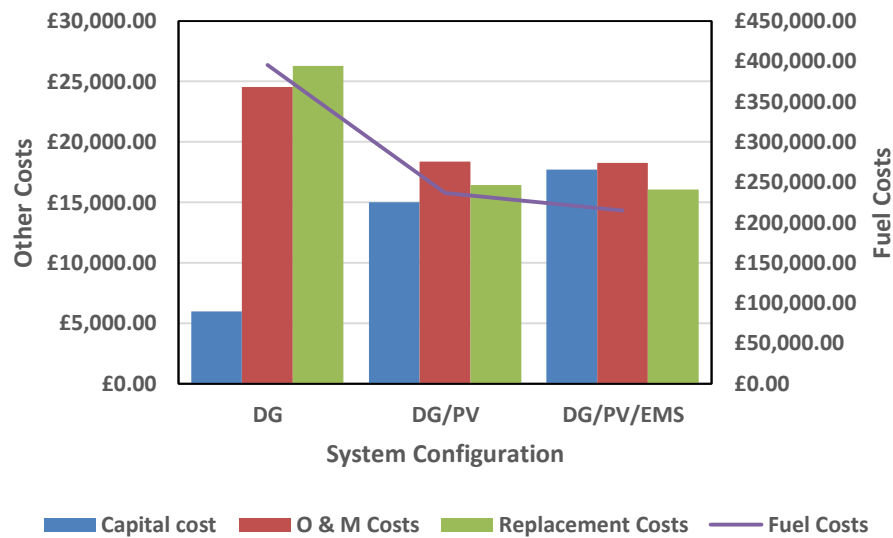


Figure 6-9 Cost comparison of DG, DG/PV, and DG/PV/EMS configurations

In Table 6-7 and Figure 6-9, DG/PV/EMS's performance in comparison to DG/PV and DG shows a reduction of about 8% and 44% in lifetime expenditure costs, and a reduction of about 9% and 46% in CO₂ emissions respectively. To determine which of the two hybrid configurations should be selected, additional parameters need to be considered. With the data from Table 6-7 and the assumptions from Table 6-4, we can begin to evaluate the financial forecasting tools highlighted in Section 6.2.2.

6.4.4 Techno-economic analysis

Two separate analyses are done with and without the impact of CB. In the fully automated scenario, the lifetime net income is calculated by multiplying the income or savings by the DF. The savings/incoming is derived by subtracting the lifetime expenditure of each configuration from the base case. Additional income is generated through energy exported to the grid, the off-peak TLC and DLC rates, and savings from the replacement costs. For instance, in 20 years, the generator in the base case would need to be changed 12 times based on a 15000-hour lifespan whereas in both hybrid configurations, the generator will be changed 7 times. The difference between the replacements costs is considered a savings. The sum of the average monthly PV output supplied to the load and the monthly PV export, multiplied by DF, makes up the lifetime PV generation.

This is then added to the DG lifetime output to make up the lifetime net electricity generation.

To investigate the impact of CB, we run the same analysis mentioned above but include assumptions made in Section 6.3.2.4. A 14% override is applied to the automated TOU schedules (heater and water pump). Out of the 14%, we ensure that 80% of the load occurs from 12am to 5am (precondition) and the other 20% occurs from 5pm to midnight (rebound). We also do the same for the DLC plug loads. By using Equations (6-2) to (6-7), Table 6-8 provides the comparison outcome using the financial forecasting tools.

Table 6-8 Techno-economic analysis of DG/PG and DG/PV/IEMS configurations

				Performance		
	DG/PV	DG/PV/IEMS Without CB	DG/PV/IEMS with CB	DG/PV	DG/PV/IEMS without CB	DG/PV/IEMS with CB
Lifetime expenditure	£271,403.21	£249,276.55	£252,537.35		✓	
Lifetime net income	£177,886.29	£204,887.85	£200,782.84		✓	
Lifetime net electricity generation	161,939 kWh	129,771 kWh	137,834 kWh	✓		
PP	1.74 years	1.73 years	1.76 years		✓	
NPV	£162,386.29	£187,187.85	£183,082.84		✓	
LCCA	£109,016.92	£62,088.69	£69,454.52		✓	
CBR	1.61	1.30	1.35		✓	
LCoE	£1.68/kWh	£1.92/kWh	£1.83/kWh	✓		
CO2 Emissions	391,703 kg CO2e	355,881 kg CO2e	361,130 kg CO2e		✓	

According to Table 6-8, the DG/PV/EMS without CB outperforms DG/PV and DG/PV/IEMS with CB in seven out of the nine categories. Although the DG/PV/EMS with CB outperforms DG/PV in six out of nine categories, it is outperformed by DG/PV/EMS without CB confirming that overrides can reduce the value and objectives of the DLC program [75]. Comparing these results to Table 6-2, three indicators highlighted in red show unfavourable results. The

LCCA, the CBR and the LCoE all suggest that both configurations should be rejected. The LCCA should be less than zero, the CBR should be less than 1 and the LCoE should be close to or comparable with the average residential electricity rate. Currently, the annual cap of household energy prices in the UK is £2,500, or the equivalent of 34.0p/kWh [76]. This is way less than the calculated prices of £1.68/kWh and £1.92/kWh. An explanation is that as Figure 6-9 suggests, DGs used as the primary supply source are always associated with high maintenance and fuel costs. In the end, the DG/PV/EMS without CB is the most favourable investment to make, earning the most income and savings with the least expenditure and CO₂ emissions over the project's entire lifetime.

Conclusion

An Integrated Energy Management System (IEMS) that combines Supply-Demand Side Management was proposed with the objective of managing and increasing photovoltaic (PV) usage in the grid. This method integrates solar energy forecasting (SEF), time-of-use (TOU), direct load control (DLC) and generator control (GC), prioritising load and energy management before calling up additional power supply. The performance of the IEMS was investigated by performing a techno-economic analysis on three off-grid configurations: diesel generator (DG), DG/PV and a fully automated DG/PV/IEMS. Using a DG as the primary supply, its associated lifetime fuel cost meant that forecasting parameters like Life Cycle Cost Analysis (LCCA), Cost-Benefit Ratio (CBR) and Levelized Cost of Energy (LCoE) made all three configurations undesirable to implement. However, the DG/PV/IEMS configuration showed a reduced life expenditure cost and CO₂ emissions of about 44% and 46% respectively when compared to DG, plus 8% and 9% respectively when compared to DG/PV. More significantly, it increased the PV usage from the DG/PV by over 113%. To analyse the impact of customer behaviour (CB), a 14% load override, 80% preconditioning and 20% rebounding were factored in based on averages given in literature. The DG/PV/IEMS without CB outperformed the DG/PV/IEMS with CB proving that overrides reduce the value of the DLC program.

The impact of using DG as the primary supply source can be negated if it is replaced with another source that uses less fuel. The versatility of the IEMS framework allows for swapping both supply/demand side responses, giving way to different system permutations. For instance, swapping the DG with a utility supply to simulate an on-grid power supply, substituting wind energy forecasting for solar, and critical peak pricing for TOU and so on. Therefore, a recommendation for future research will be to explore various combinations of supply/demand responses and evaluate their impact on solar energy integration. Exporting excess PV to an islanded or neighbourhood grid, or using dummy loads, is critical to reducing energy wastage in off-grid applications without storage. A comparison can be made between the options mentioned above or including a battery energy storage system (BESS) can be made. In addition, the impact of the IEMS can be further investigated by increasing the share of interruptible and schedulable loads. For this study, the schedulable load and interruptible DLC loads represent about 43% and 12% of the daily total load respectively. It would be worth investigating what acceptable percentages of the load can be scheduled or interrupted without causing severe impact to the user.

Another potential research is testing the IEMS rule-based algorithms using more granular historic data involving seconds rather than hourly. This will allow the corrective controls to be expressed in real-time. Finally, a CB distribution can be modelled based on historical data and applied to the analysis instead of using averages. This will present a more accurate analysis of customer's response to demand signals.

Credit Author Statement

Tolulope O. Falope: Conceptualization, Methodology, Software, Data curation, Formal Analysis, Visualization, Writing – original draft, **Lao Liyun:** Writing – review & editing, Supervision, **Da Huo:** Writing – review & editing, Supervision. , **Boyu Kuang:** Writing – review & editing

References

- [1] Shirvanova T, Zakhidova S. GLOBAL ENERGY SECURITY: PROBLEMS AND PROSPECTS. 70 Th International Scientific Conference on Economic and Social Development-Baku 2021:25–6.
- [2] Maghami MR, Mutambara AGO. Challenges associated with Hybrid Energy Systems: An artificial intelligence solution. *Energy Reports* 2023;9:924–40. <https://doi.org/10.1016/J.EGYR.2022.11.195>.
- [3] Hou L, Dong J, Herrera OE, Mérida W. Energy management for solar-hydrogen microgrids with vehicle-to-grid and power-to-gas transactions. *Int J Hydrogen Energy* 2023;48:2013–29. <https://doi.org/10.1016/J.IJHYDENE.2022.09.238>.
- [4] Gaonwe TP, Kusakana K, Hohne PA. A review of solar and air-source renewable water heating systems, under the energy management scheme. *Energy Reports* 2022;8:1–10. <https://doi.org/10.1016/J.EGYR.2022.10.065>.
- [5] Shakeri M, Shayestegan M, Reza SMS, Yahya I, Bais B, Akhtaruzzaman M, et al. Implementation of a novel home energy management system (HEMS) architecture with solar photovoltaic system as supplementary source. *Renew Energy* 2018;125:108–20. <https://doi.org/10.1016/J.RENENE.2018.01.114>.
- [6] Falope T, Lao L, Hanak D, Huo D. Hybrid energy system integration and management for solar energy: A review. *Energy Conversion and Management: X* 2024;21:100527. <https://doi.org/10.1016/J.ECMX.2024.100527>.
- [7] Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews* 2020;124:109792. <https://doi.org/10.1016/J.RSER.2020.109792>.
- [8] Zhang G, Yang D, Galanis G, Androulakis E. Solar forecasting with hourly updated numerical weather prediction. *Renewable and Sustainable Energy Reviews* 2022;154:111768. <https://doi.org/10.1016/J.RSER.2021.111768>.

- [9] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renewable and Sustainable Energy Reviews* 2022;161:112348. <https://doi.org/10.1016/J.RSER.2022.112348>.
- [10] Karunanithi K, Saravanan S, Prabakar BR, Kannan S, Thangaraj C. Integration of Demand and Supply Side Management strategies in Generation Expansion Planning. *Renewable and Sustainable Energy Reviews* 2017;73:966–82. <https://doi.org/10.1016/J.RSER.2017.01.017>.
- [11] Aoun A, Ibrahim H, Ghandour M, Ilinca A. Supply Side Management vs. Demand Side Management of a Residential Microgrid Equipped with an Electric Vehicle in a Dual Tariff Scheme. *Energies* 2019, Vol 12, Page 4351 2019;12:4351. <https://doi.org/10.3390/EN12224351>.
- [12] Pascual J, Arcos-Aviles D, Ursúa A, Sanchis P, Marroyo L. Energy management for an electro-thermal renewable-based residential microgrid with energy balance forecasting and demand side management. *Appl Energy* 2021;295:117062. <https://doi.org/10.1016/J.APENERGY.2021.117062>.
- [13] Aghajani GR, Shayanfar HA, Shayeghi H. Demand side management in a smart micro-grid in the presence of renewable generation and demand response. *Energy* 2017;126:622–37. <https://doi.org/10.1016/J.ENERGY.2017.03.051>.
- [14] Suman GK, Guerrero JM, Roy OP. Optimisation of solar/wind/bio-generator/diesel/battery based microgrids for rural areas: A PSO-GWO approach. *Sustain Cities Soc* 2021;67:102723. <https://doi.org/10.1016/J.SCS.2021.102723>.
- [15] Abedrabboh K, Al-Fagih L. Applications of mechanism design in market-based demand-side management: A review. *Renewable and Sustainable Energy Reviews* 2023;171:113016. <https://doi.org/10.1016/J.RSER.2022.113016>.
- [16] Hohne PA, Kusakana K, Numbi BP. Model validation and economic dispatch of a dual axis pv tracking system connected to energy storage with grid

connection: A case of a healthcare institution in South Africa. *J Energy Storage* 2020;32:101986. <https://doi.org/10.1016/J.EST.2020.101986>.

[17] Moradi-Sarvestani S, Jooshaki M, Fotuhi-Firuzabad M, Lehtonen M. Incorporating direct load control demand response into active distribution system planning. *Appl Energy* 2023;339:120897. <https://doi.org/10.1016/J.APENERGY.2023.120897>.

[18] Rocky Mountain Institute. Demand Response: An Introduction 2006:1–46.

[19] Pal S, Thayer B, Barrett E, Studarus K. Estimating the impacts of direct load control programs using GridPIQ, a web-based screening tool. 2017 IEEE Conference on Technologies for Sustainability, SusTech 2017 2018;2018-January:1–8. <https://doi.org/10.1109/SUSTECH.2017.8333505>.

[20] Swetha MT, Nikhil S, Lakshmi PM, Mounika T, Girish Kumar G, Harsha Y, et al. Solar Based Battery Charging System Through IOT. *International Journal of Research in Engineering and Science (IJRES)* ISSN n.d.;10:1225–32.

[21] Diouf B, Pode R. Potential of lithium-ion batteries in renewable energy. *Renew Energy* 2015;76:375–80. <https://doi.org/10.1016/J.RENENE.2014.11.058>.

[22] Hariprakash B, Martha SK, Ambalavanan S, Gaffoor SA, Shukla AK. Comparative study of lead-acid batteries for photovoltaic stand-alone lighting systems. *J Appl Electrochem* 2008;38:77–82. <https://doi.org/10.1007/S10800-007-9403-4/FIGURES/5>.

[23] Srikranjana M, Junlakorn S, Hoonchareon N. How an integration of home energy management and battery system affects the economic benefits of residential pv system owners in Thailand. *Sustainability (Switzerland)* 2021;13:1–20. <https://doi.org/10.3390/SU13052681>.

[24] IRENA. Off-grid Renewable Energy Statistics 2022. Abu Dhabi: 2022.

[25] Ritchie H, Roser M, Rosado P. Energy. *Our World in Data* 2022. <https://ourworldindata.org/energy> (accessed July 4, 2023).

- [26] Javed H, Muqeet HA, Shehzad M, Jamil M, Khan AA, Guerrero JM. Optimal energy management of a campus microgrid considering financial and economic analysis with demand response strategies. *Energies (Basel)* 2021;14. <https://doi.org/10.3390/EN14248501>.
- [27] Ouédraogo S, Faggianelli GA, Notton G, Duchaud JL, Voyant C. Impact of electricity tariffs and energy management strategies on PV/Battery microgrid performances. *Renew Energy* 2022;199:816–25. <https://doi.org/10.1016/J.RENENE.2022.09.042>.
- [28] Cupples S, Abtahi A, Madureira A, Quadrado J. Techno-Economic Feasibility Analysis and Optimal Design of Hybrid Renewable Energy Systems Coupled with Energy Storage. *Lecture Notes in Networks and Systems* 2022;419 LNNS:786–96. https://doi.org/10.1007/978-3-030-96299-9_74.
- [29] Mohandes B, Wahbah M, Moursi MS el, El-Fouly THM. Renewable energy management system: Optimum design and hourly dispatch. *IEEE Trans Sustain Energy* 2021;12:1615–28. <https://doi.org/10.1109/TSTE.2021.3058252>.
- [30] George-Williams H, Wade N, Carpenter RN. A probabilistic framework for the techno-economic assessment of smart energy hubs for electric vehicle charging. *Renewable and Sustainable Energy Reviews* 2022;162. <https://doi.org/10.1016/J.RSER.2022.112386>.
- [31] Mahmud K, Hossain MJ, Ravishankar J. Peak-Load Management in Commercial Systems with Electric Vehicles. *IEEE Syst J* 2019;13:1872–82. <https://doi.org/10.1109/JSYST.2018.2850887>.
- [32] Anusha U, Jagadeesh S, Rao KV, Umavani M. Energy Management of an Off-Grid and Grid Connected Hybrid Renewable Energy Source Micro-Grid System for Commercial Load. *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference* 2022. <https://doi.org/10.1109/MYSURUCON55714.2022.9972673>.

- [33] Ali S, Malik TN, Raza A. Risk-Averse Home Energy Management System. IEEE Access 2020;8:91779–98. <https://doi.org/10.1109/ACCESS.2020.2994462>.
- [34] Kichou S, Skandalos N, Wolf P. Energy performance enhancement of a research centre based on solar potential analysis and energy management. Energy 2019;183:1195–210. <https://doi.org/10.1016/J.ENERGY.2019.07.036>.
- [35] Ozden T, Kesler S, Okumus HI. A fuzzy logic embedded energy management software with multi-agent system for a stand-alone PV power plant. International Journal of Environmental Science and Technology 2019;16:5197–204. <https://doi.org/10.1007/S13762-018-2144-6/FIGURES/8>.
- [36] Rochd A, Benazzouz A, Ait Abdelmoula I, Raihani A, Ghennioui A, Naimi Z, et al. Design and implementation of an AI-based & IoT-enabled Home Energy Management System: A case study in Benguerir — Morocco. Energy Reports 2021;7:699–719. <https://doi.org/10.1016/J.EGYR.2021.07.084>.
- [37] Harajli H, Kabakian V, El-Baba J, Diab A, Nassab C. Commercial-scale hybrid solar photovoltaic - diesel systems in select Arab countries with weak grids: An integrated appraisal. Energy Policy 2020;137. <https://doi.org/10.1016/J.ENPOL.2019.111190>.
- [38] Dinh HT, Yun J, Kim DM, Lee KH, Kim D. A Home Energy Management System with Renewable Energy and Energy Storage Utilizing Main Grid and Electricity Selling. IEEE Access 2020;8:49436–50. <https://doi.org/10.1109/ACCESS.2020.2979189>.
- [39] Silva VA, Aoki AR, Lambert-Torres G. Optimal day-ahead scheduling of microgrids with battery energy storage system. Energies (Basel) 2020;13. <https://doi.org/10.3390/EN13195188>.
- [40] Ochoa T, Gil E, Angulo A. Efficient Bidding of a PV Power Plant with Energy Storage Participating in Day-Ahead and Real-Time Markets Using Artificial Neural Networks. IEEE Power and Energy Society General Meeting 2022;2022-July. <https://doi.org/10.1109/PESGM48719.2022.9916732>.

- [41] Sharda S, Sharma K, Singh M. A real-time automated scheduling algorithm with PV integration for smart home prosumers. *Journal of Building Engineering* 2021;44:102828. <https://doi.org/10.1016/J.JOBE.2021.102828>.
- [42] Tabrizi AR. Energy Management System Optimization for Grid Connected Microgrids in Presence of Energy Storage; *Energy Management System Optimization for Grid Connected Microgrids in Presence of Energy Storage* 2022. <https://doi.org/10.1109/HORA55278.2022.9800062>.
- [43] Rammelkamp K, Schröder S, Kubitz S, Vogt DS, Frohmann S, Hansen PB, et al. Low-level LIBS and Raman data fusion in the context of in situ Mars exploration. *Journal of Raman Spectroscopy* 2020;51:1682–701. <https://doi.org/10.1002/JRS.5615>.
- [44] Falope TO, Lao L, Hanak D. A Unique Three-Step Weather Data Approach in Solar Energy Prediction Using Machine Learning. *Sustainable Energy Solutions for a Post-COVID Recovery towards a Better Future: Part VII*, Bangkok: International Conference on Applied Energy ; 2021, p. 1–7. <https://doi.org/https://doi.org/10.46855/energy-proceedings-9863>.
- [45] Abdelhady S. Performance and cost evaluation of solar dish power plant: sensitivity analysis of levelized cost of electricity (LCOE) and net present value (NPV). *Renew Energy* 2021;168:332–42. <https://doi.org/10.1016/J.RENENE.2020.12.074>.
- [46] Mellichamp DA. Internal rate of return: Good and bad features, and a new way of interpreting the historic measure. *Comput Chem Eng* 2017;106:396–406. <https://doi.org/10.1016/J.COMPCHEMENG.2017.06.005>.
- [47] Acaroğlu H, García Márquez FP. A life-cycle cost analysis of High Voltage Direct Current utilization for solar energy systems: The case study in Turkey. *J Clean Prod* 2022;360:132128. <https://doi.org/10.1016/J.JCLEPRO.2022.132128>.
- [48] Sajid MU, Bicer Y. Comparative life cycle cost analysis of various solar energy-based integrated systems for self-sufficient greenhouses. *Sustain Prod Consum* 2021;27:141–56. <https://doi.org/10.1016/J.SPC.2020.10.025>.

- [49] Androsch H, Ramasubramanian V, Foster R, Kabir H, Orozco R, Stadter H, et al. Mini-Grid Design. 2nd ed. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH Nigerian, Energy Support Programme (NESP); 2017.
- [50] Eze F, Ogola J, Kivindu R, Egbo M, Obi C. Technical and economic feasibility assessment of hybrid renewable energy system at Kenyan institutional building: A case study. *Sustainable Energy Technologies and Assessments* 2022;51. <https://doi.org/10.1016/J.SETA.2021.101939>.
- [51] Gerber DL, Ghatpande OA, Nazir M, Heredia WGB, Feng W, Brown RE. Energy and power quality measurement for electrical distribution in AC and DC microgrid buildings. *Appl Energy* 2022;308:118308. <https://doi.org/10.1016/J.APENERGY.2021.118308>.
- [52] Elma O, Taşcıkaraoğlu A, Tahir İnce A, Selamoğulları US. Implementation of a dynamic energy management system using real time pricing and local renewable energy generation forecasts. *Energy* 2017;134:206–20. <https://doi.org/10.1016/J.ENERGY.2017.06.011>.
- [53] Minhas DM, Khalid RR, Frey G. Load control for supply-demand balancing under Renewable Energy forecasting. 2017 IEEE 2nd International Conference on Direct Current Microgrids, ICDCM 2017 2017:365–70. <https://doi.org/10.1109/ICDCM.2017.8001071>.
- [54] International Monetary Fund. World Economic Outlook Database. International Monetary Fund 2023. <https://www.imf.org/en/Publications/WEO/weo-database/2023/April> (accessed April 24, 2023).
- [55] Elmorshedy MF, Elkadeem MR, Kotb KM, Taha IBM, El-Nemr MK, Kandeal AW, et al. Feasibility study and performance analysis of microgrid with 100% hybrid renewables for a real agricultural irrigation application. *Sustainable Energy Technologies and Assessments* 2022;53. <https://doi.org/10.1016/J.SETA.2022.102746>.

- [56] Grant Thornton. Renewable energy discount rate survey results - 2017. 2018.
- [57] Danish SN, Al-Ansary H, El-Leathy A, Ba-Abbad M, Khan SUD, Rizvi A, et al. Experimental and techno-economic analysis of two innovative solar thermal receiver designs for a point focus solar Fresnel collector. *Energy* 2022;261. <https://doi.org/10.1016/J.ENERGY.2022.125035>.
- [58] SUNSTORE. 10kw Solar Kit. n.d. <https://www.sunstore.co.uk/product/10kw-solar-kit/> (accessed May 1, 2023).
- [59] Cymru Solar. Commercial Solar System Cost. Cymru Solar 2023. <https://southwalessolarpanels.co.uk/commercial-solar-system-cost/> (accessed May 1, 2023).
- [60] Trevithick D. 10kW Solar System UK: Price, How Much Does It Produce, How Big, and More. UpLevel Green 2023. https://uplevelgreen.com/uk/solar-systems/10-kw-solar-panels-uk/?utm_content=cmp-true (accessed May 1, 2023).
- [61] Ramasamy V, Feldman D, Desai J, Margolis R. U.S. Solar Photovoltaic System and Energy Storage Cost Benchmarks: Q1 2021 2021.
- [62] Babajide A, Brito MC. Solar PV systems to eliminate or reduce the use of diesel generators at no additional cost: A case study of Lagos, Nigeria. *Renew Energy* 2021;172:209–18. <https://doi.org/10.1016/J.RENENE.2021.02.088>.
- [63] Das BK, Zaman F. Performance analysis of a PV/Diesel hybrid system for a remote area in Bangladesh: Effects of dispatch strategies, batteries, and generator selection. *Energy* 2019;169:263–76. <https://doi.org/10.1016/J.ENERGY.2018.12.014>.
- [64] Kumar P, Pal N, Sharma H. Techno-economic analysis of solar photovoltaic/diesel generator hybrid system using different energy storage technologies for isolated islands of India. *J Energy Storage* 2021;41:102965. <https://doi.org/10.1016/J.EST.2021.102965>.

- [65] Ghenai C, Bettayeb M. Modelling and performance analysis of a stand-alone hybrid solar PV/Fuel Cell/Diesel Generator power system for university building. *Energy* 2019;171:180–9. <https://doi.org/10.1016/J.ENERGY.2019.01.019>.
- [66] Ismail MS, Moghavvemi M, Mahlia TMI. Techno-economic analysis of an optimized photovoltaic and diesel generator hybrid power system for remote houses in a tropical climate. *Energy Convers Manag* 2013;69:163–73. <https://doi.org/10.1016/J.ENCONMAN.2013.02.005>.
- [67] Greaves Cotton Engineering Company. Diesel Generator Set Specifications Sheet- 7.5 kVA n.d. https://www.greavescotton.com/php/media/css_images/brochure/specsheet-7.5.pdf (accessed May 19, 2023).
- [68] Department for Energy Security and Net Zero. Greenhouse gas reporting: conversion factors 2023. GOVUK 2023. <https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2023> (accessed July 5, 2023).
- [69] Department for Energy Security and Net Zero. Weekly road fuel prices. National Statistics 2013. <https://www.gov.uk/government/statistics/weekly-road-fuel-prices> (accessed June 4, 2023).
- [70] Elexon Limited. System Sell & System Buy Prices. BMRS 2023. <https://www.bmreports.com/bmrs/?q=balancing/systemsellbuyprices/historic> (accessed June 4, 2023).
- [71] Kaczmarek J, Jones B, Chermak J. Determinants of Demand Response Program Participation: Contingent Valuation Evidence from a Smart Thermostat Program. *Energies* 2022, Vol 15, Page 590 2022;15:590. <https://doi.org/10.3390/EN15020590>.
- [72] Xu X, Chen C fei, Zhu X, Hu Q. Promoting acceptance of direct load control programs in the United States: Financial incentive versus control option. *Energy* 2018;147:1278–87. <https://doi.org/10.1016/J.ENERGY.2018.01.028>.

- [73] Laitos VM, Bargiotas D, Daskalopulu A, Arvanitidis AI, Tsoukalas LH. An Incentive-Based Implementation of Demand Side Management in Power Systems. *Energies* 2021, Vol 14, Page 7994 2021;14:7994. <https://doi.org/10.3390/EN14237994>.
- [74] Sarran L, Gunay HB, O'Brien W, Hviid CA, Rode C. A data-driven study of thermostat overrides during demand response events. *Energy Policy* 2021;153:112290. <https://doi.org/10.1016/J.ENPOL.2021.112290>.
- [75] Wildstein PJ, Craig MT, Vaishnav P. Participant overrides can halve the reliability value of direct load control programs. *Energy Build* 2023;299:113606. <https://doi.org/10.1016/J.ENBUILD.2023.113606>.
- [76] Department for Energy Security & Net Zero. Energy Price Guarantee. GOVUK 2023. <https://www.gov.uk/government/publications/energy-bills-support/energy-bills-support-factsheet-8-september-2022> (accessed July 17, 2023).

7 Discussions, conclusions, and recommendations

7.1 Discussions

The current state-of-the art for RES integration is categorized into various control techniques, integration approaches, and power system modelling and optimization algorithms including integration of microgrids, distributed energy resources (DER) and energy storage systems (ESS) [1]. In considering the above methods, it was observed from the findings of the literature review in Chapter 2 that there was no clear framework for solar energy management and integration into today's modern grid. Solar energy integration has a lot of connected parts but there is no overarching framework to link all the components. To further explore this gap, all the above-mentioned techniques had to be analysed.

Power system modelling and optimization in terms of RES integration have been highly researched and is a well-established research theme. Optimization problems usually include set objective(s) that must be optimised considering pre-determined constraints. The goal is to derive an optimal solution that minimizes or maximizes the objective function. Example of optimization algorithms include Mixed Integer Linear Programming [2], Genetic Algorithm [3], Particle Swarm Optimization [4], Sequential Quadratic Programming [5], Rolling-horizon optimization [6] and so on. These algorithms describe the mathematics of achieving supply-demand balance, focusing less on the structure of a framework. Power or energy system modelling use software like PLEXOS or H2RES to solve unit commitment or energy dispatch problems through multi-objective decision optimizations. Most software are black boxes and not openly sourced, with no description of their architecture or methodologies.

Control techniques connected to the management of the power system come close and are broadly separated into demand side management (DSM) and supply side management (SSM) [7]. SSM and DSM are both critical in integration of variable renewable energy (VRE) sources [8] and are critical in multi-energy systems (MES) and building applications [9]. While several literatures investigate DSM [10,11] and SSM [12,13] separately, others propose that controlling both

supply and demand is critical for VRE integration [14] and must be considered in designing reliable and financially viable systems [15]. For example, Dranka et al. [16] propose an integrated co-optimization model for demand (load shedding and shifting) and RES supply coordination based on linear programming. Karunanithi et al. [17] consider energy efficiency and peak shifting as DSM strategies and upgrading transmission distribution lines and increasing thermal power plants process efficiency as SSM strategies. Further study of the literature revealed an additional component implicitly linked to SSM and DSM. This component is energy forecasting. Narvaez et al. states that for a grid with available PV plants, predicting the available solar energy is essential [18]. Accurate solar energy forecasting (SEF) reduces uncertainty [14,19] caused by solar intermittency. In addition, supply and demand forecasting helps to reduce uncertainty and effectively matches energy supply and demand [20]. Therefore, one of the outputs of the literature review is identifying the key components of a solar-based integrated energy management system (IEMS) which are SEF, SSM and DSM.

In reviewing integration approaches, integration energy systems (IES) and energy management systems (EMS) were regularly referenced in literature. IES is defined as a system that integrates multiple energy and power sources using advance energy conversion technologies to efficiently improve energy use, increase system flexibility and reduce carbon emissions [21,22]. An EMS monitors, optimizes, and controls the consumers' load, power generation and transmission [23]. The traditional EMS was coined when the power supply was largely homogenous. However, this definition is becoming obsolete as the integration of the internet-of-things (IoT), storage and electric vehicles have increased the grid complexity [23]. A new definition of the EMS that covers today's complex grid is essential. The concept of IEMS was introduced and is thought to refer to today's multi-source, multi-energy and multi control grid. Wang et al. [24] defines an IEMS as a platform that integrates energy dispatch and control for multiple energy sources. Looking at this definition, it is relatively similar to the definitions of IES and EMS. It is no wonder that all three definitions are sometimes used interchangeably and often defined as an optimization problem. However, the findings in Chapter 2 makes a distinction between the three by

giving a clear definition of an IEMS in the context of today's advanced grid. A summary of findings from Chapter 2 are listed below:

- EMS, IES, optimization modelling, and IEMS are sometimes used interchangeably even though they may refer to different aspects of integration.
- IEMS is a system that manages multiple energy sources by leveraging on advancement in technology and communication to integrate both predictive and real-time controls, and initiate supply and demand responses to balance the load and power supply in the grid.
- Uncertainty in the power system can be reduced if integrated renewable energy sources (RES) can be predicted.
- Advances in grid technology and communication, a diverse energy mix, and increased interconnections, mean that the new EMS must be able to interact with both the supply and demand sides of the grid. The grid has become smarter so the EMS must also become smarter.
- Renewable smart hybrid mini-grids (RSHMG) possess the requisite technology and infrastructure for solar energy integration.
- The solar-based IEMS structure or framework has several possible combinations because of the different type of SEF approaches, SSM, and DSM responses. The choice of a combination of approaches/responses is often dictated by constraints and objectives unique to the case study or location.
- Forecasting is not a 100% science; real time load scheduling needs to be adopted to adjust the prediction errors.

By assessing the outputs of the literature review, it is clear to see that Objective 1 has been achieved. Different solar-based integrated configurations have been assessed and a comprehensive definition of IEMS has been developed. In addition, the different components of a solar-based IEMS have been identified

leading to the research question of how the components would interact within the IEMS.

Chapter 3 and 4 pick up on the findings from Chapter 2 by focusing on SEF as one of the IEMS components, highlighting the fact that uncertainty can be reduced if solar energy is accurately predicted. The motivation to use a numerical weather prediction (NWP)-based forecasting approach is because NWP incorporates real-time weather data to continuously refine the accuracy of predictions while adapting to changing atmospheric conditions. This dynamic capability makes NWP effective in improving forecasting accuracy. With access to three roof-top and a utility-scale PV plant, half-hourly PV meter readings, an on-site weather station, a local weather station, a three-step forecasting approach was developed. The foundation of this method is based on the following:

- Effectively leveraging the connections between heterogeneous data sources from multiple sites is crucial for accurate forecasting [25].
- Multivariate approaches perform better than univariate [26] with multiple data sources providing richer than information from any one individual data source [27].
- An on-site weather station improves the quality of weather data [28].

The three-step forecasting architecture proposed in Chapter 3 offers a simplistic way of forecasting, by combining multiple raw data sources using Low-level Data Fusion (LLDF). The reduced root mean square error (RMSE) of this approach compared to approaches using a univariate data source supports the achievement of Objective 2. Furthermore, this approach was tested and validated using data from a 1MW solar farm and three rooftop solar arrays (8 kW, 10.5 kW and 15 kW), showing its ability to be used for the entire spectrum of distributed generation customers. What this approach lacks in complexity, it makes up in accuracy by using data preprocessing (DP) to improve the quality of input data into the forecasting model. Six DP categories and sixteen DP methods for statistical data-driven PV forecasting architecture were considered in Chapter 4. Though not an exhaustive list of DP techniques, they proved to be very effective when deployed in data driven statistical forecasting models like the proposed

method in Chapter 3. In this case, the model's accuracy was improved by reducing the RMSE by 75%.

Chapter 5 focuses on direct load control (DLC), which is a sub-category under DSM (identified as one of the key concepts of an IEMS in Chapter 2). A case study analysis was performed to determine the effectiveness of DLC in reducing energy wastage and consumption in small power plug-in loads. DLC is also one of the most implemented demand response (DR) programs designed to manage small commercial or domestic customer loads [29]. In a comparison of DR programs in Chapter 2, sub-section 5, DLC is the only incentive-based DR program that can be pre-notified or controlled in real time. This characteristic fits in perfectly with the IEMS definition of Chapter 2, highlighting the importance of both predictive and real-time controls. In this case study, the aim was to identify periods where energy could be turned off when not in use. With help from Honeywell's Connected Power Solutions, two test sites in Cranfield were selected, a robotics laboratory and a café. These two locations with completely different load consumption profiles and plug-in devices were selected to validate DLC. A baseline was established before the DLC was implemented to determine peak (operating times) and off-peak times of the devices/appliances. Using the Baseline comparison method, results show that the DLC reduced energy wastage by 47%. This is significant because the DLC's ability to reduce unwanted and unused energy is a critical role in the IEMS's ability to match demand to supply. Thus, Objective 3 was achieved.

The literature review properly identifies SEF, DSM, and SSM as crucial components of the IEMS but raises a research question of their interconnectedness within the IEMS. By reviewing the definition of the IEMS, any potential framework must have a predictive and real-time component. In a summary table of different IEMS configurations in Chapter 2, sub-section 8, different studies use different combinations of SEF, SSM and DSM for solar energy integration. After careful review of the literature, a novel framework was created in Chapter 6:

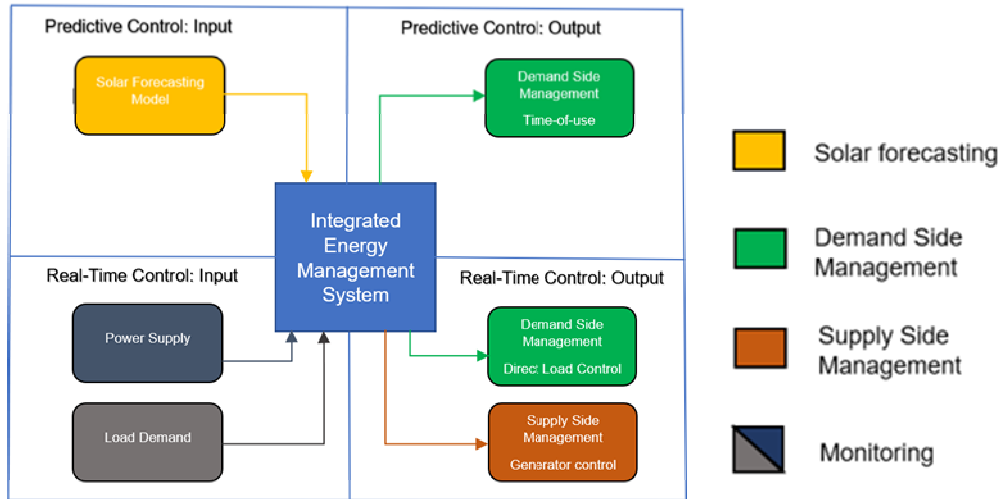


Figure 7-1 Proposed solar-based IEMS framework

The framework in Figure 7-1 was designed in such a way that it allows different SEF, DSM and SSM techniques to be investigated while providing predictive and real-time control of the power system which is an indispensable part of efficient VRE integration. In other words, it allows the user to substitute any combination of SEF, DSM, SSM in the highlighted blocks. This framework can work in both off-grid and on-grid applications in the presence or absence of a utility grid. Apart from designing this framework, the novelty lies in the fact that there has not been a unique combination of SEF, time-of-use (TOU) and DLC as DSM, and generator control (GC) SSM within the IEMS in any of the available literature. In this framework, the three-step forecasting architecture developed and validated in Chapter 3 and 4 is used as the SEF model. The DLC component validated in Chapter 5 is also used. The choice of TOU is based on findings from the literature review in Chapter 2 and is considered as one of the main pricing strategies used in the power market [30]. The choice of a diesel generator (DG) as a SSM technique is also informed by literature review in Chapter 2. Typically, in areas where grid extension is not economically feasible, stand-alone RES and DGs have been deployed to meet load demand [31]. A DG can be automatically or manually controlled and may be used to meet peak loads in cases of BESS depletion, rapid changes in load, or RES intermittency [32]. DGs are widely used for off-grid solutions, and therefore offer versatility as they can be used for both off-grid and on-grid applications. Therefore, the choice of DGs with GC was

adopted. However, DGs can be switched out for a utility source in on-grid applications, which shows the versatility of the framework.

The combined use of override, preconditioning, and rebound rates to model customer behaviour is also a novel concept. Preconditioning (before DLC event) and rebounding (after DLC event) are customer behaviours that result in an increase in the customer's load consumption before and/or after a DLC event or during a TOU peak period to compensate for load curtailment, higher usage costs or threat of penalties [33]. To assess its technical, economical, and environmental impact, the IEMS is deployed along with a DG/PV configuration in a residential house. The DG/PV/IEMS configuration showed a reduced life expenditure cost and CO₂ emissions of about 44% and 46% respectively when compared to a solely DG source, plus 8% and 9% respectively when compared to the DG/PV configuration. In addition, the DG/PV/IEMS configuration increased PV usage by over 113% up from the DG/PV configuration, thus achieving Objective 4. Table 7-1 presents an overview of each objective's novelty, scientific soundness, and contributions. This table combines the findings from each chapter, showing how they fit together to accomplish the main aim of the study:

Table 7-1 Summary of novelty and findings

Objective	Novelty	Scientific soundness (Methods/validation)
To assess different IEMS configurations and determine how individual components increase the reliability of solar-based power systems.	Chapter 2 presents a definition and a comprehensive literature review of IEMS, including a review of other integrated approaches for solar-based hybrid systems.	A thorough review of each component and their deployment in real-world solar integration problems was conducted and materials sourced from relevant scientific databases. A second review was simultaneously conducted on different integrated approaches using components from the first review.
To develop a method that uses LLDF in combining weather parameters from an on-site weather station and a local weather station to	Chapter 3 and 4 presents evidence that no SEF methods have been identified that use LLDF specifically to combine data for both an on-	Forecasting PV generation based on solar radiation is an already validated technique. LLDF has been affirmed to also increase credible

improve solar energy forecasting.	<p>site and local weather station, while identifying forecasting predictors using solar radiation correlation.</p> <p>A framework for classifying DP methods for data-driven statistical PV forecasting models is developed in Chapter 4.</p>	<p>features in data. After this model was developed using data from 1 MW farm, it was validated using three separate rooftop arrays.</p>
To ascertain the efficacy of DLC in reducing energy wastage and consumption, thereby affirming its role as a feasible demand response strategy for optimizing load to align with available supply.	<p>The case study in Chapter 5 highlight's the use of DLC to manage power for small plug-in loads at the outlet level. The solution used here uniquely combines Mira Mesh RF and BACnet/IP to integrate DLC into a buildings energy management system (BEMS).</p>	<p>Tests were conducted on the ability of the system to measure half-hourly energy consumption and respond to scheduling commands in two different sites. During the scheduling phase, load schedule times were advised by users and implemented via the software. Users confirmed equipment were turned off and on as scheduled.</p>
To develop a comprehensive descriptor for assessing the impact of the proposed IEMS on solar hybrid applications, encompassing technical, economic, and environmental dimensions.	<p>Chapter 6 builds on from Chapter 2 - 5 and a novel IEMS framework is created integrating SEF from Chapter 3 and 4, TOU and DLC (validated from Chapter 5) as DSM, GC as SSM.</p> <p>In addition, customer behaviour was uniquely modelled using averages supported by literature for override, preconditioning, and rebound rates.</p>	<p>After the literature review identified the possible components of the IEMS, an IEMS framework was proposed. Review of the state-of-the-art revealed that there is no IEMS framework with a unique combination of SEF, TOU, DLC and GC. To validate the performance, the IEMS framework was tested using a 1-year residential load. The IEMS proved to increase solar penetration by over 113%.</p>

7.2 Conclusion

In today's complex and advanced grid, IEMS must cater to multi-source, multi-control and multi-load requirements. The role of RES like solar energy will only increase if we are to achieve sustainable goals of reducing GHG emissions. Solar energy's variable nature means supply will not always match demand. Therefore, to sustainably increase the share of solar energy in the grid without compromising grid stability, the aim of this research was directed towards maintaining the demand-supply energy balance. This aim was achieved by meeting the objectives highlighted in Section 1.3.

To achieve these objectives, a literature review was first conducted to better understand the state-of-the-art of IEMS, EMS, IES and other integrated approaches, and how they balance demand and supply of intermittent RES systems. The review revealed that today's solar-based IEMS has three critical components: SEF, DSM and SSM. This is the first time these components have been explicitly linked in a framework and an understanding of their interconnectedness outlined. On the bases of this foundation, the study continued to examine different approaches of SEF, DSM and SSM, through model development, case study and deterministic analysis, culminating in the proposed IEMS framework that integrates SEF, TOU and DLC as examples of DSM responses, and GC as an example of SSM response. With a general IEMS framework identified, users can experiment with different permutations of SEF, DSM and SSM responses to see which unique combination gives them the best result based on their unique objectives.

7.2.1 Three-step forecasting model development

In Chapter 3 and 4, a three-step forecasting architecture is proposed. This model combines weather parameters from both an on-site weather station and a nearby local weather station through LLDF. This proposed method is heavily data dependent, so the type of data and the source is critical to its accuracy and performance. One of the criteria used for selecting weather data was results obtained from the cross-validation of correlated coefficients using Pearson Correlation Coefficient (PCC) analysis. For weather variables measured by both

stations, obtaining approximately the same correlated coefficients to Solar radiation, was a key data reliability indicator given the relative closeness of both stations to each other. This cross-validation technique helps mitigate against and identify data biases that may exist in one of the weather stations if the coefficients are considerably different. Other considerations for data selection included proximity to the PV plants to minimize spatial variability in weather conditions. The temporal resolution (between 15 mins and one hour) of weather variables showed good enough data granularity to capture weather characteristics. Furthermore, the accessibility and availability of historic weather and PV data, plus the robust range of recorded weather variables were critical in selecting the data sources. In both Chapter 3 and 4, data preprocessing was used to clean and improve data quality and reduce the effects of outliers that may affect forecasting performance.

Combining useful data from more than one source increased the number of relevant weather parameters as predictors in the forecasting model and improved its performance. Another function of PCC analysis is to identify parameters with moderate to strong positive correlations to solar radiation. In addition, raw meteorological data was preprocessed using the Rescaling (RM) and List-wise deletion (LWDM) methods, which proved to be effective in increasing the accuracy of the model. A deterministic analysis was performed by training and testing the architecture using MATLAB's Regression Learner App with data from a 1 MW solar farm and validated using data from three rooftop solar arrays (8 kW, 10.5 kW and 15 kW), achieving root RMSE values of 0.0859, 0.0984, 0.1425 and 0.0885 respectively. The low RMSE values, a measure of the predicted PV to actual PV generation, proves that the model can be adopted for different PV plant sizes.

To improve the accuracy of the model, sixteen DP models were investigated in Chapter 4. Out of these sixteen methods, only linear interpolation, RM, outlier detection, data smoothing, Pearson correlation coefficient, day/time/month decomposition, seasonal decomposition, principal component analysis, and holdout validation had a positive impact on model accuracy. In a

comparative study using two and a half years of weather and PV data, results show that the accuracy of the three-step forecasting model discussed in Chapter 3, with LWDM, PCC, RM, and LLDF methods, was improved by 75% when the additional DP methods mentioned above were included in the preprocessing. DP methods should always be considered especially for heavily dependent data-driven statistical models like the proposed model in Chapter 3.

7.2.2 Direct load control: A case study

DLC is a viable DR tool that can help reduce wastage and energy consumption. In a case study highlighted in Chapter 5, DLC reduced energy consumption by 47% across two locations in Cranfield university. The ability to remotely schedule devices/appliances on and off is a critical DSM response that can either reduce consumption to match supply or increase consumption during favourable or off-peak tariff periods.

7.2.3 Deterministic analysis of the integrated energy management framework

An IEMS that combines supply-demand side management was developed with the objective of managing and increasing PV usage in the grid. This method integrates SEF, TOU, DLC and GC, prioritising load and energy management before calling up additional power supply. The performance of the IEMS was investigated by performing a deterministic analysis on the techno-economic outputs of three off-grid configurations: DG, DG/PV and a fully automated DG/PV/IEMS. Using a DG as the primary supply, its associated lifetime fuel cost meant that forecasting parameters like Life Cycle Cost Analysis (LCCA), Cost-Benefit Ratio (CBR) and Levelized Cost of Energy (LCoE) made all three configurations undesirable to implement. However, the DG/PV/IEMS configuration showed a reduced life expenditure cost and CO₂ emissions of about 44% and 31% respectively on DG, plus 8% and 2% respectively on DG/PV. More significantly, it increased the PV usage from the DG/PV by over 113%. To analyse the impact of customer behaviour (CB), a 14% load override, 80% preconditioning and 20% rebounding were factored in based on averages deduced from

literature. The DG/PV/IEMS without CB outperformed the DG/PV/IEMS with CB proving that overrides reduce the value of the DLC program.

7.2.4 Key contributions to knowledge

This research focused on first identifying the knowledge gap, developing an SEF model, validating DLC through a case study, and performing deterministic analysis on both the SEF model and the proposed IEMS framework. The key contributions to knowledge are:

1. The results and findings of the literature review provide an understanding of the IEMS state-of-the-art including methods and best practices for optimizing RES, knowledge gaps, limitations, and recommendations. A comprehensive definition of IEMS was derived along with guidance on possible IEMS framework (Chapter 2).
2. The proposed three-step forecasting approach contributes to advancing the field of SEF by exploring the potential benefits of LLDF. Such findings could be vital for optimizing solar power generation and reducing dependency on non-RES (Chapter 3).
3. The creation of a DP framework for data-driven PV statistical forecasting models will assist researchers in better scoping additional DP methods for a more systematic analysis of raw data to improve forecasting accuracy (Chapter 4).
4. The outcomes and conclusions of the case study contribute to the state-of-the-art in DLC as a DR tool to manage and reduce energy consumption. Different DLC configurations (including software and hardware) and architectures were analysed and profiled. Results for this market-ready product were obtained to gain a fair assessment of the technology (Chapter 5).
5. The deterministic method on the techno-economic analysis of the proposed IEMS framework is expected to create a holistic understanding

of various aspects, including technical feasibility, economic viability, and environmental sustainability. This IEMS framework can help investigate different combinations of DR and SSM controls, determining the right combination for each unique scenario and objective (Chapter 6).

The growing consensus from many relevant stakeholders is to prioritize RES like solar energy over traditional fossil fuels. This study is significant and gives regulators and operators the tools to choose the most appropriate strategy for planning and grid management. Consumers on the demand-side could have a more dynamic role in managing their energy and making better informed decisions, with predictive and real-time controls. Each part of this study can be evaluated independently to help advance the discussion of solar energy integration for a sustainable future.

7.2.5 Research limitations

The few identified research limitations are:

- The weather and PV data available from the on-site weather station and PV meter respectively, were limited to a few years. A richer data source would have perhaps given additional insights into the forecasting results. In addition, the availability of more historical data helps to identify trends as well as data anomalies.
- The review of DP methods in Chapter 4 is not exhaustive. More DP methods need to be scoped and properly classified to give a more comprehensive picture of preprocessing techniques used for statistical PV forecasting models.
- The effectiveness of TOU was not independently validated with a case study. The author relied on findings in the literature. Just like the other components in this study, a case study on TOU would have supported and validated conclusions from the literature.

- A residential building was not used in the DLC case study to validate its performance for residential applications. In addition, energy consumption drivers were not visible in the data due to the shortened trial duration. The number of corresponding days in each of the phases (baseline, scheduling, and optimization) was also irregular.
- The most granular data available for use in validating the IEMS was half-hourly data. This meant that a real-time response was not possible necessitating a change to a corrective-response.

7.3 Recommendations for future research

The proposed areas for future research include:

- The proposed SEF model can be improved upon by investigating ways of replacing missing or inconsistent data. The model works better when the raw data before the data fusion is as complete and authentic as possible. Other methods like Multivariate Imputation by Chained Equations (MICE), forward and backward fill, along with domain-specific methods can be investigated.
- Data covering a longer period should be obtained from older in-situ weather stations and verifiable open-source government databases. These data sources do not have to be limited to the UK and can be domiciled in other countries.
- DP literature has been well researched and developed. More literature reviews should be carried out to allow for validation and eventual classification of additional DP methods for a more comprehensive DP framework. This can include DP methods like Wavelet and EMD based decompositions, Singular spectrum analysis for denoising, Metaheuristic optimization-based feature selections, filter-based correction and so on.
- TOU can be validated by using a real-life case study with the use of smart meters or advanced metering infrastructure (AMI). For example,

the impact of the Shape Shifters TOU tariff created by Octopus energy on electricity prices and savings can be assessed.

- In the IEMS framework, various combinations of supply/demand responses can be investigated to determine their impact on solar energy integration. For example, TOU can be replaced with real-time pricing (RTP) to determine which response returns higher energy savings for the customer.
- The IEMS framework can be tested using more granular PV and load consumption data preferably in real-time. This will allow a proper validation of all PV forecasting, load, and generator scheduling algorithms.
- In the deterministic analysis performed in Chapter 6, the schedulable load and interruptible DLC loads represented about 43% and 12% of the daily total load respectively. Possible future research would be to investigate what acceptable percentages of the load can be scheduled or interrupted without causing severe impact to the user.
- The impact of BESS either as a SSM response or independently on its own has not been thoroughly investigated within the IEMS framework. In addition, the DG could be replaced with a utility to determine the impact of the IEMS framework on on-grid applications.
- For future research, Honeywell's Connected Power Solution can be extended to residential buildings and a comparison made to its performance on commercial buildings. In addition, energy consumption drivers like occupancy and seasonality should be considered in determining the overall energy used in the baselining and optimization phases. This will give a more accurate result of the average daily energy consumed and saved, while reducing the standard deviation due to the sporadic nature of the power usage data. The baseline, scheduling and optimization timelines should be increased to better identify usage patterns and possible intra-variable relations.

References

- [1] Ikram M, Aslam M, Aurangzeb K, Ahmed S, Marwat SNK, Haider SI, et al. Integrating renewable energy sources for optimal demand-side management using decentralized multi-agent control. *Sustainable Energy, Grids and Networks* 2023;36:101193. <https://doi.org/10.1016/J.SEGAN.2023.101193>.
- [2] Guo Y, Xiang Y. IESP: An integrated energy system planning platform. *Energy Reports* 2023;9:242–8. <https://doi.org/10.1016/J.EGYR.2023.04.280>.
- [3] Su H, Chi L, Zio E, Li Z, Fan L, Yang Z, et al. An integrated, systematic data-driven supply-demand side management method for smart integrated energy systems. *Energy* 2021;235. <https://doi.org/10.1016/J.ENERGY.2021.121416>.
- [4] Pawar P, TarunKumar M, Vittal K. P. An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement* 2020;152:107187. <https://doi.org/10.1016/J.MEASUREMENT.2019.107187>.
- [5] Ben Hmida J, Chambers T, Lee J. Solving constrained optimal power flow with renewables using hybrid modified imperialist competitive algorithm and sequential quadratic programming. *Electric Power Systems Research* 2019;177:105989. <https://doi.org/10.1016/J.EPSR.2019.105989>.
- [6] Craparo EM, Sprague JG. Integrated supply- and demand-side energy management for expeditionary environmental control. *Appl Energy* 2019;233–234:352–66. <https://doi.org/10.1016/J.APENERGY.2018.09.220>.
- [7] Jabir HJ, Teh J, Ishak D, Abunima H. Impacts of demand-side management on electrical power systems: A review. *Energies (Basel)* 2018;11. <https://doi.org/10.3390/en11051050>.
- [8] Dranka GG, Ferreira P. Review and assessment of the different categories of demand response potentials. *Energy* 2019;179:280–94. <https://doi.org/10.1016/J.ENERGY.2019.05.009>.

- [9] Luo XJ, Fong KF. Development of integrated demand and supply side management strategy of multi-energy system for residential building application. *Appl Energy* 2019;242:570–87. <https://doi.org/10.1016/J.APENERGY.2019.03.149>.
- [10] Von Hayn A, Moske J, Aselmann P, Weigold M. Demand side management in the cooling supply of brewing processes. *Procedia CIRP* 2022;105:146–51. <https://doi.org/10.1016/J.PROCIR.2022.02.025>.
- [11] Noor S, Guo M, Van Dam KH, Shah N, Wang X. Energy Demand Side Management with supply constraints: Game theoretic Approach. *Energy Procedia* 2018;145:368–73. <https://doi.org/10.1016/J.EGYPRO.2018.04.066>.
- [12] Tayerani Charmchi AS, Ifaei P, Yoo CK. Smart supply-side management of optimal hydro reservoirs using the water/energy nexus concept: A hydropower pinch analysis. *Appl Energy* 2021;281:116136. <https://doi.org/10.1016/J.APENERGY.2020.116136>.
- [13] Peng C, Peng T, Liu Y, Geissdoerfer M, Evans S, Tang R. Industrial Internet of Things enabled supply-side energy modelling for refined energy management in aluminium extrusions manufacturing. *J Clean Prod* 2021;301:126882. <https://doi.org/10.1016/J.JCLEPRO.2021.126882>.
- [14] Mahmood D, Javaid N, Ahmed G, Khan S, Monteiro V. A review on optimization strategies integrating renewable energy sources focusing uncertainty factor – Paving path to eco-friendly smart cities. *Sustainable Computing: Informatics and Systems* 2021;30:100559. <https://doi.org/10.1016/J.SUSCOM.2021.100559>.
- [15] Karunanithi K, Saravanan S, Prabakar BR, Kannan S, Thangaraj C. Integration of Demand and Supply Side Management strategies in Generation Expansion Planning. *Renewable and Sustainable Energy Reviews* 2017;73:966–82. <https://doi.org/10.1016/J.RSER.2017.01.017>.

- [16] Dranka GG, Ferreira P, Vaz AIF. Integrating supply and demand-side management in renewable-based energy systems. *Energy* 2021;232:120978. <https://doi.org/10.1016/J.ENERGY.2021.120978>.
- [17] Karunanithi K, Saravanan S, Prabakar BR, Kannan S, Thangaraj C. Integration of Demand and Supply Side Management strategies in Generation Expansion Planning. *Renewable and Sustainable Energy Reviews* 2017;73:966–82. <https://doi.org/10.1016/J.RSER.2017.01.017>.
- [18] Narvaez G, Giraldo LF, Bressan M, Pantoja A. Machine learning for site-adaptation and solar radiation forecasting. *Renew Energy* 2021;167:333–42. <https://doi.org/10.1016/J.RENENE.2020.11.089>.
- [19] Cox S, Xu K. Integration of Large-Scale Renewable Energy in the Bulk Power System: Good Practices from International Experiences. National Renewable Energy Laboratory 2020.
- [20] Zhou K, Chu Y, Hu R. Energy supply-demand interaction model integrating uncertainty forecasting and peer-to-peer energy trading. *Energy* 2023;285:129436. <https://doi.org/10.1016/J.ENERGY.2023.129436>.
- [21] Wang Y, Qin Y, Ma Z, Wang Y, Li Y. Operation optimisation of integrated energy systems based on cooperative game with hydrogen energy storage systems. *Int J Hydrogen Energy* 2023;48:37335–54. <https://doi.org/10.1016/J.IJHYDENE.2023.06.170>.
- [22] Hengrui M, Wang B, Wang H, Wang S, Fuqi M, Wang X, et al. An effective planning approach for integrated energy systems considering equipment operating characteristics. *Heliyon* 2023;9:e21409. <https://doi.org/10.1016/J.HELİYON.2023.E21409>.
- [23] Tajjour S, Singh Chandel S. A comprehensive review on sustainable energy management systems for optimal operation of future-generation of solar microgrids. *Sustainable Energy Technologies and Assessments* 2023;58:103377. <https://doi.org/10.1016/J.SETA.2023.103377>.

- [24] Wang X, Li Y, Wang C, Lin H, Li M. Research on Integrated Energy Management System Based on Microservice Architecture. Proceedings - 2021 International Conference on Power System Technology: Carbon Neutrality and New Type of Power System, POWERCON 2021 2021:401–6. <https://doi.org/10.1109/POWERCON53785.2021.9697774>.
- [25] Almaghrabi S, Rana M, Hamilton M, Saiedur Rahaman M. Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks. Information Fusion 2024;104:102180. <https://doi.org/10.1016/J.INFFUS.2023.102180>.
- [26] Scabbia G, Sanfilippo A, Bachour D, Perez-Astudillo D. Exogenous Parameters in Solar Forecasting. Conference Record of the IEEE Photovoltaic Specialists Conference 2020;2020-June:0894–6. <https://doi.org/10.1109/PVSC45281.2020.9300800>.
- [27] Wang B, Che J, Wang B, Feng S. A Solar Power Prediction Using Support Vector Machines Based on Multi-source Data Fusion. 2018 International Conference on Power System Technology, POWERCON 2018 - Proceedings 2019:4573–7. <https://doi.org/10.1109/POWERCON.2018.8601672>.
- [28] Moreno-Muñoz A editor. Large scale grid integration of renewable energy sources. Institution of Engineering and Technology; 2017.
- [29] Miri M, McPherson M. Demand response programs: Comparing price signals and direct load control. Energy 2024;288:129673. <https://doi.org/10.1016/J.ENERGY.2023.129673>.
- [30] Zhao X, Gao W, Qian F, Ge J. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system. Energy 2021;229:120538. <https://doi.org/10.1016/J.ENERGY.2021.120538>.
- [31] Come Zebra EI, van der Windt HJ, Nhumaio G, Faaij APC. A review of hybrid renewable energy systems in mini-grids for off-grid electrification in developing countries. Renewable and Sustainable Energy Reviews 2021;144:111036. <https://doi.org/10.1016/J.RSER.2021.111036>.

[32] Suman GK, Guerrero JM, Roy OP. Optimisation of solar/wind/bio-generator/diesel/battery based microgrids for rural areas: A PSO-GWO approach. *Sustain Cities Soc* 2021;67:102723. <https://doi.org/10.1016/J.SCS.2021.102723>.

[33] Pal S, Thayer B, Barrett E, Studarus K. Estimating the impacts of direct load control programs using GridPIQ, a web-based screening tool. 2017 IEEE Conference on Technologies for Sustainability, SusTech 2017 2018;2018-January:1–8. <https://doi.org/10.1109/SUSTECH.2017.8333505>.

APPENDICES

Appendix A A Three-step weather data approach in solar energy prediction using machine learning

A.1 Performance of forecasting architecture

Table A-1 Performance comparison of regression models in RLA For Scenario 1

	Train				Test			
Model	MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²
LINEAR REGRESSION								
Linear	0.0772	0.0155	0.1244	0.76	0.0562	0.0093	0.0967	0.71
Interactions Linear	0.0722	0.0143	0.1197	0.78	0.0955	0.0145	0.1206	0.55
Robust Linear	0.0735	0.0185	0.1360	0.72	0.0688	0.0123	0.1110	0.62
Stepwise Linear	0.0721	0.0143	0.1198	0.78	0.0920	0.0141	0.1188	0.57
REGRESSION TREES								
Fine Tree	0.0412	0.0062	0.0787	0.91	0.0486	0.0120	0.1095	0.63
Medium Tree	0.0541	0.0098	0.0990	0.85	0.0458	0.0100	0.1002	0.69
Coarse Tree	0.0599	0.0117	0.1082	0.82	0.0442	0.0091	0.0955	0.72
SUPPORT VECTOR MACHINES								
Linear SVM	0.0711	0.0158	0.1256	0.76	0.0673	0.0109	0.1046	0.66
Quadratic SVM	0.0678	0.0138	0.1174	0.79	0.0870	0.0149	0.1221	0.54
Cubic SVM	0.1181	0.0279	0.1671	0.57	0.1294	0.0469	0.2165	-0.44
Fine Gaussian	0.0603	0.0113	0.1065	0.83	0.0744	0.0134	0.1157	0.59
Medium Gaussian	0.0648	0.0128	0.1133	0.80	0.0672	0.0114	0.1070	0.65
Coarse Gaussian	0.0692	0.0140	0.1185	0.79	0.0478	0.0081	0.0901	0.75
GAUSSIAN PROCESS REGRESSION								
Rational Quadratic	0.0068	0.0002	0.0147	1.00	0.0613	0.0124	0.1115	0.62
Squared Exponential	0.0625	0.0117	0.108	0.82	0.1008	0.0239	0.1547	0.26
Matern 5/2	0.0608	0.0112	0.1057	0.83	0.0997	0.0230	0.1518	0.29
Exponential	3.03e-4	1.22e-6	0.0011	1.00	0.0744	0.0166	0.1288	0.49
ENSEMBLES OF TREES								
Boosted Trees	0.0615	0.0118	0.1087	0.82	0.0441	0.0084	0.0917	0.74
Bagged Trees	0.0502	0.0081	0.0898	0.88	0.0531	0.0093	0.0964	0.71
NEURAL NETWORK								
Narrow Neural	0.0643	0.0124	0.1112	0.81	0.0780	0.0128	0.1133	0.61
Medium Neural	0.0616	0.0116	0.1078	0.82	0.1031	0.0234	0.1529	0.28
Wide Neural	0.0596	0.0109	0.1045	0.83	0.1224	0.0327	0.1808	-0.01
Bilayered Neural	0.0621	0.0119	0.1089	0.82	0.0556	0.0108	0.1040	0.67

Trilayered Neural	0.0611	0.0115	0.1070	0.82	0.0957	0.0202	0.1422	0.38
-------------------	--------	--------	--------	------	--------	--------	--------	------

Table A-2 Performance comparison of regression models in RLA For Scenario 2

	Train				Test			
Model	MAE	MSE	RMSE	R ²	MAE	MSE	RMSE	R ²
LINEAR REGRESSION								
Linear	0.0780	0.0147	0.1213	0.78	0.0607	0.0010	0.0100	0.69
Interactions Linear	0.0739	0.0141	0.1189	0.78	0.0559	0.0097	0.0984	0.70
Robust Linear	0.0740	0.0167	0.1302	0.74	0.0465	0.0010	0.0998	0.69
Stepwise Linear	0.0739	0.0141	0.1189	0.78	0.0559	0.0097	0.0984	0.70
REGRESSION TREES								
Fine Tree	0.0429	0.0066	0.0811	0.90	0.0556	0.0145	0.1204	0.55
Medium Tree	0.0562	0.0103	0.1017	0.84	0.0505	0.0119	0.1092	0.63
Coarse Tree	0.0638	0.0126	0.1122	0.81	0.0483	0.0108	0.1037	0.67
SUPPORT VECTOR MACHINES								
Linear SVM	0.0737	0.0150	0.1226	0.77	0.0518	0.0097	0.0985	0.70
Quadratic SVM	0.0713	0.0145	0.1203	0.78	0.0500	0.0095	0.0975	0.71
Cubic SVM	0.0660	0.0133	0.1155	0.80	0.0487	0.0099	0.0997	0.69
Fine Gaussian	0.0611	0.0113	0.1064	0.83	0.0559	0.0108	0.1038	0.67
Medium Gaussian	0.0664	0.0129	0.1136	0.80	0.0497	0.0098	0.0991	0.70
Coarse Gaussian	0.0717	0.0142	0.1192	0.78	0.0503	0.0092	0.0999	0.72
GAUSSIAN PROCESS REGRESSION								
Rational Quadratic	0.0020	2.577e-05	0.0051	1.00	0.0597	0.0146	0.1209	0.55
Squared Exponential	0.0639	0.0123	0.1110	0.81	0.0495	0.0100	0.1001	0.69
Matern 5/2	0.0576	0.0104	0.1018	0.84	0.0498	0.0103	0.1015	0.68
Exponential	3.20e-4	3.81e-6	0.0020	1.00	0.0530	0.0118	0.1088	0.64
ENSEMBLES OF TREES								
Boosted Trees	0.0663	0.0128	0.1133	0.80	0.0465	0.0097	0.0987	0.70
Bagged Trees	0.0573	0.0094	0.0968	0.86	0.0522	0.0101	0.1007	0.69
NEURAL NETWORK								
Narrow Neural	0.0652	0.0128	0.1130	0.80	0.0489	0.0097	0.0985	0.70
Medium Neural	0.0643	0.0124	0.1113	0.81	0.0517	0.0103	0.1015	0.68
Wide Neural	0.0617	0.0115	0.1073	0.82	0.0523	0.0106	0.1031	0.67
Bilayered Neural	0.0630	0.0122	0.1105	0.81	0.0496	0.0105	0.1023	0.68
Trilayered Neural	0.0626	0.0120	0.1096	0.82	0.0507	0.0106	0.1030	0.67

A.2 Scenario prediction comparisons using Trapezoidal rule

Table A-3 Trapezoidal rule calculation of areas under the curves

Hours	Actual	Scenario 1	Scenario 2	Scenario 3
0	0	-0.06196	0.001677	0.005502
1	0	-0.05747	0.001677	0.00539
2	0	-0.0543	-0.00131	0.009166
3	0	-0.05265	-0.00066	0.01572
4	0.005156	-0.05272	0.008427	0.020084
5	0.044569	-0.06125	0.011387	0.045674
6	0.123169	-0.08926	0.143359	0.110453
7	0.326946	-0.09722	0.274514	0.269908
8	0.444135	0.045696	0.383478	0.500645
9	0.577094	0.252907	0.604034	0.641404
10	0.692264	0.532067	0.641365	0.753875
11	0.726599	0.724883	0.373679	0.696112
12	0.7689	0.731178	0.406016	0.679695
13	0.570539	0.692602	0.642644	0.685248
14	0.303218	0.380112	0.400898	0.333426
15	0.234184	0.092457	0.161005	0.099877
16	0.147901	0.065957	0.108772	0.079057
17	0.004814	0.036532	0.037098	0.034738
18	0	-0.00446	0.033382	0.013739
19	0	-0.01649	0.03967	0.019668
20	0	-0.02783	0.040755	0.025797
21	0	-0.03587	0.050936	0.0334
22	0	-0.03719	0.060723	0.03688
23	-	-	-	-
Approximate Area	4.969	2.906	4.424	5.115

Appendix B Data preprocessing techniques for statistical PV forecasting models

Table B-1 Performance results of applying separate data preprocessing methods to training and testing data sets

Data cleaning									
Test	LWDM	LIM	DO	FO	DS	RMSE	Algorithm		
1	✓					37.13	BOT		
2	✓		✓			37.10	BOT		
3	✓			✓		37.22	BOT		
4	✓				✓	40.36	BAT		
5	✓		✓		✓	40.43	BAT		
6	✓			✓	✓	40.28	BAT		
7		✓				42.16	LR		
8		✓	✓			37.08	BOT		
9		✓		✓		37.19	BOT		
10		✓			✓	40.38	BAT		
11		✓	✓		✓	40.31	BAT		
12		✓		✓	✓	40.34	BAT		
Data scaling									
	LWDM	LIM	DO	FO	ZS	RM	DS	RMSE	Algorithm
*13	✓				✓			107.5	MT
*14	✓					✓		108.6	MT
*15	✓		✓		✓			107.7	CT
*16	✓		✓			✓		108.6	MT/CT
*17	✓			✓	✓			107.5	MT
*18	✓			✓		✓		108.6	MT
19	✓				✓			0.4010	BAT
20	✓					✓		0.0927	BOT
21	✓		✓		✓			0.4035	BAT
22	✓		✓			✓		0.1257	LR

23	✓			✓	✓			0.3990	BAT
24	✓			✓		✓		0.0927	BOT
25	✓					✓	✓	0.0989	BAT

Note: BAT = Bagged Trees, BOT = Boosted Trees, CT = Coarse Tree, DO = Delete Outliers, DS = Data Smoothing, FO = Fill Outliers, LIM = Linear Interpolation Method, LR = Linear Regression, LWDM = List-wise Deletion Method, MT = Medium Tree, RM = Rescaling Method, RMSE = Root Mean Square Error, ZS = Z-Score

*# represent trials that used LIM and RM for the test data set while using other data processing methods for the training data set

Table B-2 Performance results of data preprocessing methods with different Pearsons Correlation Coefficients

PCC > 0.5									
Data integration and data reduction									
1	LIM	SRC	RM	ZS	Cranf	Bed	Cranf + Bed	RMSE	Algorithm
2	✓	✓	✓		✓			0.1039	TNN
3	✓	✓	✓			✓		0.1216	MNN
4	✓	✓	✓				✓	0.0965	BNN
5	✓	✓		✓			✓	0.4084	MNN
Data decomposition – Day, Time , Month, Daylight-free									
	LIM	SRC	RM	Cranf + Bed		DTM	DF	RMSE	Algorithm
6	✓	✓	✓	✓		✓		0.0856	BNN
7	✓	✓	✓	✓		✓	✓	0.1627	MGPR
8	✓	✓	✓	✓			✓	0.1744	SEGPR
Data decomposition – Seasons									
	LIM	SRC	RM	Cranf + Bed	DTM	Season	RMSE	Algorithm	
9	✓	✓	✓	✓		Spring	0.1161	SEGPR	
10	✓	✓	✓	✓	✓	Spring	0.0945	SEGPR	
11	✓	✓	✓	✓	✓	Summer	0.0850	MGPR	
12	✓	✓	✓	✓	✓	Autumn	0.0756	BAT	
13	✓	✓	✓	✓	✓	Winter	0.0811	QSVM	
PCC > 0.4									
Data integration and data reduction									
	LIM	SRC	RM	Cranf + Bed	NC	RMSE	Algorithm		

14	✓	✓	✓	✓		0.0818	TNN			
15	✓	✓	✓	✓	✓	0.0832	MNN			
Data partitioning										
	LIM	SRC	RM	Cranf + Bed	NC	DTM	Validation	RMSE	Algorithm	
16	✓	✓	✓	✓	✓	✓	Cross : 5	0.0854	BNN	
17	✓	✓	✓	✓	✓	✓	Cross: 10	0.0851	TNN	
18	✓	✓	✓	✓	✓	✓	Cross: 15	0.0840	BNN	
19	✓	✓	✓	✓	✓	✓	Cross: 20	0.0834	BNN	
20	✓	✓	✓	✓	✓	✓	Cross: 40	0.0852	TNN	
21	✓	✓	✓	✓	✓	✓	Holdout: 10	0.0840	MNN	
22	✓	✓	✓	✓	✓	✓	Holdout: 20	0.0836	WNN	
23	✓	✓	✓	✓	✓	✓	Holdout: 30	0.0817	BNN	
24	✓	✓	✓	✓	✓	✓	Holdout: 40	0.0779	BNN	
Data reduction - PCA										
	LIM	SRC	RM	Cranf + Bed	NC	DTM	PCA	Validation	RMSE	Algorithm
25	✓	✓	✓	✓	✓	✓	10	Resubstituion	0.0860	MNN
26	✓	✓	✓	✓	✓	✓	10	Holdout: 40	0.0840	BNN
27	✓	✓	✓	✓	✓	✓	8	Holdout: 40	0.0818	TNN
28	✓	✓	✓	✓	✓	✓	6	Holdout: 40	0.0873	TNN
29	✓	✓	✓	✓	✓	✓	4	Holdout: 40	0.1011	BNN
Data decomposition – Seasons										
	LIM	SRC	RM	Cranf + Bed	NC	DTM	Season	Validation	RMSE	Algorithm
30	✓	✓	✓	✓	✓	✓	Spring	Resubstituion	0.0983	EGPR
31	✓	✓	✓	✓	✓	✓	Summer	Resubstituion	0.0879	SEGPR
32	✓	✓	✓	✓	✓	✓	Autumn	Resubstituion	0.0649	NNN
33	✓	✓	✓	✓	✓	✓	Winter	Resubstituion	0.0691	EGPR
34	✓	✓	✓	✓	✓	✓	Spring	Holdout: 40	0.0876	EGPR
35	✓	✓	✓	✓	✓	✓	Summer	Holdout: 40	0.0892	MGPR
36	✓	✓	✓	✓	✓	✓	Autumn	Holdout: 40	0.0688	EGPR
37	✓	✓	✓	✓	✓	✓	Winter	Holdout: 40	0.0667	NNN

Overall performance										
	LIM	SRC	RM	Cranf + Bed	NC	DTM	PCA	Validation	RMSE	Algorithm
38	✓	✓	✓	✓	✓	✓	8	Resubstituion	0.0813	BAT
39	✓	✓	✓	✓	✓	✓	8	Holdout: 40	0.0837	TNN
	LIM	SRC	RM	Cranf + Bed	DTM	Season		RMSE		Algorithm
40	✓	✓	✓	✓		Spring		0.1161		SEGPR
41	✓	✓	✓	✓	✓	Spring		0.0945		SEGPR

Note: BAT = Bagged Trees, Bed = Bedford, BNN = Bilayered Neural Network, Cranf. = Cranfield, Cross : # = Cross-validation : Number of folds, DF = Daylight-free, DTM = Day-Time-Month, Holdout = Holdout Validation : Percent hold out, LIM = Linear Interpolation Method, LWDM = List-wise Deletion Method, MGPR = Matern 5/2 Gaussian Process Regression, MNN = Medium Neural Network, NC = Negative Correlation, NNN = Narrow Neural Network, PCA = Principal Component Analysis, QSVM = Quadratic Support Vector Machine, RM = Rescaling Method, RMSE = Root Mean Square Error, SEGPR = Squared Exponential Gaussian Process Regression, SRC = Solar Radiation Correlation, Trilayered Neural Network, WNN = Wide Neural Network, ZS = Z-Score

Table B-3 Performance results of data preprocessing methods

Data cleaning, data scaling, data reduction, and data integration											
	LWDM	LIM	FO	DS	RM	SRC	Cranf	Bed	Cranf + Bed	RMSE	Algorithm
1	✓		✓		✓	✓	✓			0.1042	NNN
2	✓		✓		✓	✓		✓		0.1073	MNN
3	✓		✓		✓	✓			✓	0.0928	BNN
4	✓		✓	✓	✓	✓	✓			0.0604	WNN
5	✓		✓	✓	✓	✓		✓		0.0688	FGSVM
6	✓		✓	✓	✓	✓			✓	0.0549	FGSVM
7		✓	✓		✓	✓	✓			0.1037	BNN
8		✓	✓		✓	✓		✓		0.1049	TNN
9		✓	✓		✓	✓			✓	0.0946	MNN
10		✓	✓	✓	✓	✓			✓	0.0547	WNN
Data scaling											
	LIM	FO	DS	ZS	RM	SRC	PVC	Cranf +Bed	RMSE		Algorithm
11	✓	✓	✓	✓		✓		✓	0.2552		WNN

12	✓	✓	✓		✓	✓		✓	0.0547	WNN		
13	✓	✓	✓		✓		✓	✓	0.0543	MNN		
Data decomposition												
	LIM	FO	DS	RM	Cranf + Bed	PVC	DTM	NC	DF	RMSE	Algorithm	
14	✓	✓	✓	✓	✓	✓	✓			0.0283	FGSVM	
15	✓	✓	✓	✓	✓	✓	✓	✓		0.0268	WNN	
16	✓	✓	✓	✓	✓	✓	✓	✓	✓	0.1984	WNN	
Data decomposition – Seasons												
	LIM	FO	DS	RM	Cranf + Bed	PVC	DTM	NC	Season	RMSE	Algorithm	
17	✓	✓	✓	✓	✓	✓	✓	✓	Summer	0.0128	RQGPR	
18	✓	✓	✓	✓	✓	✓	✓	✓	Spring	0.0144	MGPR	
19	✓	✓	✓	✓	✓	✓	✓	✓	Autumn	0.0090	SEGPR	
20	✓	✓	✓	✓	✓	✓	✓	✓	Winter	0.0079	MGPR	
Data/Dimentionality reduction – PCA												
	LIM	FO	DS	RM	Cranf + Bed	PVC	DTM	NC	PCA	RMSE	Algorithm	
21	✓	✓	✓	✓	✓	✓	✓	✓	PCA 11	0.0265	WNN	
22	✓	✓	✓	✓	✓	✓	✓	✓	PCA 9	0.0264	WNN	
23	✓	✓	✓	✓	✓	✓	✓	✓	PCA 8	0.0133	WNN	
24	✓	✓	✓	✓	✓	✓	✓	✓	PCA 7	0.0276	WNN	
25	✓	✓	✓	✓	✓	✓	✓	✓	PCA 5	0.0522	FGSVM	
Data partitioning												
	LIM	FO	DS	RM	Cranf + Bed	PVC	DTM	NC	PCA	Validation	RMSE	Algorithm
26	✓	✓	✓	✓	✓	✓	✓	✓	9	Cross : 5	0.0280	MGSVM
27	✓	✓	✓	✓	✓	✓	✓	✓	9	Cross: 10	0.0272	WNN
28	✓	✓	✓	✓	✓	✓	✓	✓	9	Cross: 15	0.0265	WNN
29	✓	✓	✓	✓	✓	✓	✓	✓	9	Cross: 20	0.0268	WNN
30	✓	✓	✓	✓	✓	✓	✓	✓	9	Holdout: 10	0.0269	WNN
31	✓	✓	✓	✓	✓	✓	✓	✓	9	Holdout: 20	0.0265	WNN
32	✓	✓	✓	✓	✓	✓	✓	✓	9	Holdout: 30	0.0272	WNN
33	✓	✓	✓	✓	✓	✓	✓	✓	9	Holdout: 40	0.0263	WNN

Data decomposition								
	Seasons	Validation	PCA	RMSE	Algorithm	PCA	RMSE	Algorithm
34	Summer	Holdout: 40	9	0.0112	RQGPR	8	0.0096	RQGPR
35	Spring	Holdout: 40	9	0.0103	RQGPR	8	0.0108	RQGPR
36	Autumn	Holdout: 40	9	0.0055	MGPR	8	0.0075	MGPR
37	Winter	Holdout: 40	9	0.0064	RQGPR	8	0.0072	MGPR

Note: Bed = Bedford, BNN = Bilayered Neural Network, Cranf. = Cranfield, Cross : # = Cross-validation : Number of folds, DF = Daylight-free, DO = Delete Outliers, DS = Data Smoothing, DTM = Day-Time-Month, FGSVM = Fine Gaussian Support Vector Machine, FO = Fill Outliers, Holdout = Holdout Validation : Percent hold out, LIM = Linear Interpolation Method, LWDM = List-wise Deletion Method, MGPR = Matern 5/2 Gaussian Process Regression, MGSVM = Medium Gaussian Support Vector Machine, MNN = Medium Neural Network, NC = Negative Correlation, NNN = Narrow Neural Network, PCA = Principal Component Analysis, PVC = Photovoltaic Correlation, RM = Rescaling Method, RMSE = Root Mean Square Error, RQGPR = Rational Quadratic Gaussian Process Regression, SEGPR = Squared Exponential Gaussian Process Regression, SRC = Solar Radiation Correlation, TNN = Trilayered Neural Network, WNN = Wide Neural Network, ZS = Z-Score

Table B-4 Performance results of data preprocessing methods using only one year of data

LWMD performance									
	LWMD	SRC	PVC	RM	Cranf + Bed	NC	Validation	RMSE	Algorithm
1	✓	✓		✓	✓		Resubstitution	0.0859	BOT
2	✓		✓	✓	✓		Resubstitution	0.0864	BOT
3	✓	✓		✓	✓	✓	Resubstitution	0.0858	BOT
4	✓		✓	✓	✓	✓	Resubstitution	0.0865	BOT
5	✓	✓		✓	✓		Cross : 5	0.0859	BOT
6	✓	✓		✓	✓	✓	Cross: 10	0.0859	BOT
7	✓	✓		✓	✓	✓	Cross: 15	0.0859	BOT
8	✓	✓		✓	✓	✓	Cross: 20	0.0859	BOT
9	✓	✓		✓	✓	✓	Holdout: 10	0.0859	BOT
10	✓	✓		✓	✓	✓	Holdout: 20	0.0859	BOT
11	✓	✓		✓	✓	✓	Holdout: 30	0.0859	BOT
12	✓	✓		✓	✓	✓	Holdout: 40	0.0859	BOT
Data reduction - PCA									
	Validation		PCA	RMSE	Algorithm	PCA	RMSE	Algorithm	

13	Resubstitution	95%	0.1807	RL	7	0.0103	RL
14	Resubstitution	90%	0.1807	RL	5	0.1010	RL
15	Resubstitution	75%	0.1509	MG SVM	3	0.1023	RL
16	Resubstitution	50%	0.1509	MG SVM			

Note: Bed = Bedford, BOT = Boosted Trees, Cranf. = Cranfield, Cross : # = Cross-validation : Number of folds, Holdout = Holdout Validation : Percent hold out, LWDM = List-wise Deletion Method, NC = Negative Correlation, PCA = Principal Component Analysis, PVC = Photovoltaic Correlation, RL = Robust Linear, RM = Rescaling Method, RMSE = Root Mean Square Error, SRC = Solar Radiation Correlation

