

DOCTORAL THESIS

Energy Management
System for Single-Cell
Three-Phase Energy Router
in Residential Applications

Hossein Nourollahi Hokmabad

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
91/2025

Energy Management System for Single-Cell Three-Phase Energy Router in Residential Applications

HOSSEIN NOUROLLAHI HOKMABAD



TALLINN UNIVERSITY OF TECHNOLOGY

School of Engineering

Department of Electrical Power Engineering and Mechatronics

This dissertation was accepted for the defence of the degree 21/10/2025

Supervisor: Prof. Dmitri Vinnikov
Department of Electrical Power Engineering and Mechatronics
Tallinn University of Technology, Tallinn, Estonia

Co-supervisor: Assoc. Prof. Oleksandr Husev
Department of Industrial Electronics
Warsaw University of Technology, Warsaw, Poland

Co-supervisor: Assoc. Prof. Juri Belikov
Department of Software Science
Tallinn University of Technology, Tallinn, Estonia

Opponents: Prof. Amjad Anvari-Moghaddam
The Faculty of Engineering and Science
Department of Energy
Aalborg University, Aalborg, Denmark

Assoc. Prof. Hugo Morais
Departamento de Engenharia Electrotécnica e de Computadores
Instituto Superior Técnico, Lisbon, Portugal

Defence of the thesis: 12/12/2025, Tallinn

Declaration:

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Hossein Nourollahi Hokmabad

signature



Copyright: Hossein Nourollahi Hokmabad, 2025

ISSN 2585-6898 (publication)

ISBN 978-9916-80-415-5 (publication)

ISSN 2585-6901 (PDF)

ISBN 978-9916-80-416-2 (PDF)

DOI <https://doi.org/10.23658/taltech.91/2025>

Printed by Koopia Niini & Rauam

Nourollahi Hokmabad, H. (2025). *Energy Management System for Single-Cell Three-Phase Energy Router in Residential Applications* [TalTech Press]. <https://doi.org/10.23658/taltech.91/2025>

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
91/2025

Energiahaldussüsteem üheelemendilisele kolmefaasilisele energiaruuterile elamutes

HOSSEIN NOUROLLAHI HOKMABAD



Contents

List of Publications.....	7
Author’s Contribution to the Publications	8
Abbreviations	9
Symbols	11
1 Introduction	14
1.1 Background	14
1.2 Motivation of the Thesis.....	15
1.3 Aims, Hypothesis, and Research Tasks	15
1.4 Research Methods	16
1.5 Contributions and Disseminations.....	17
1.6 Experimental Setup and Instruments	18
1.7 Thesis Outline	19
2 Proposed Hardware and Software Platform for Home EMS.....	20
2.1 State-of-the-Art	20
2.1.1 Research Niches	22
2.1.2 Technical Niches.....	22
2.2 Proposed Home EMS	22
2.3 Summary.....	28
3 Solar Power Generation and Electricity Consumption Forecasting	29
3.1 Day Ahead Solar Power Generation Forecasting Algorithm	29
3.1.1 Physics-Based Model.....	30
3.1.2 ML Based Model	31
3.1.3 Numerical Results	33
3.2 Day Ahead Electricity Consumption Forecasting Algorithm	36
3.3 Summary.....	37
4 Developed Energy Management Method.....	38
4.1 Heuristic (Rule) Based Approach	39
4.2 Deterministic approach	39
4.3 Probabilistic Approach.....	44
4.4 Performance Comparison	49
4.4.1 Key Performance Indicators (KPIs).....	49
4.4.2 Sensitivity Analysis	53
4.5 Summary.....	55
5 Size Optimization and Cost Justification	56
5.1 Techno-Economic Model Development	56
5.1.1 Battery Degradation Model	56
5.1.2 Techno-Economic Model	58
5.1.3 Results.....	60

5.2 Single-Cell Three-Phase Solution	64
5.2.1 Single-Cell Three-Phase EnergyR Topology	64
5.2.2 Analysed Scenarios and Problem Formulation	65
5.3 Summary.....	71
6 Performance Report, and Future Direction	72
6.1 Performance Report	72
6.1.1 Day-Ahead Solar Forecasting Performance	74
6.1.2 Optimization Performance Monitoring.....	74
6.1.3 Home Battery Operation Logs.....	76
6.2 Future Direction.....	76
6.2.1 Digital Twinning.....	76
6.3 Summary.....	80
7 Conclusions	82
List of Figures	84
List of Tables	86
References	87
Acknowledgements.....	92
Abstract.....	93
Lühikokkuvõte.....	94
Appendix	95
Curriculum Vitae	189
Elulookirjeldus.....	190

List of Publications

The following is a list of author's publications, on the basis of which the thesis has been prepared:

- [Paper I] H. N. Hokmabad, O. Husev, J. Kurnitski, and J. Belikov, 'Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications', *Sustainable Energy, Grids and Networks*, vol. 38, p. 101385, 2024.
- [Paper II] H. N. Hokmabad, O. Husev, and J. Belikov, 'Day-Ahead Solar Power Forecasting Using LightGBM and Self-Attention Based Encoder-Decoder Networks', *IEEE Transactions on Sustainable Energy*, vol. 16, no. 2, pp. 866–879, 2025.
- [Paper III] N. Shabbir, K. Vassiljeva, H. Nourollahi Hokmabad, O. Husev, E. Petlenkov, and J. Belikov, 'Comparative Analysis of Machine Learning Techniques for Non-Intrusive Load Monitoring', *Electronics*, vol. 13, no. 8, 2024.
- [Paper IV] H. N. Hokmabad, T. H. Shahsavari, O. Matiushkin, T. Jalakas, O. Husev, and J. Belikov, 'Single Cell Energy Router Justification for Three Phase Near Zero Energy Buildings', in *2025 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2025, pp. 1622–1628.
- [Paper V] H. N. Hokmabad, T. H. Shahsavari, P. P. Vergara, O. Husev, and J. Belikov, 'Forecast-Driven Scenario Generation for Building Energy Management Using Stochastic Optimization', *arXiv [eess.SY]*. 2025.
- [Paper VI] H. N. Hokmabad, N. Shabbir, V. Astapov, E. Petlenkov, O. Husev, and J. Belikov, 'Feasibility Study of a DC House Connected to a Conventional AC Distribution Network', in *2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2024, pp. 1–6.
- [Paper VII] H. N. Hokmabad, O. Husev, D. Vinnikov, J. Belikov, and E. Petlenkov, 'Day-ahead PV Output Power Forecasting Utilizing Boosting Recursive LightGBM-LSTM Framework', in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2023, pp. 1–5.

Author's Contribution to the Publications

Contribution to the papers in this thesis are:

- [Paper I] Hossein Nourollahi Hokmabad was a main author of the paper. He has developed an optimization framework and optimal sizing analysis based on demand and generation considering energy storage degradation impacts. He has provided a literature analysis, calculation processes, and writing. He was responsible for submission and contact with editors during peer-reviewing rounds.
- [Paper II] Hossein Nourollahi Hokmabad was a main author of the paper. He has developed a novel day-ahead solar forecasting method based on physics-based models and machine learning techniques. He has provided a literature analysis, calculation processes, and writing. He was responsible for submission and contact with editors during peer-reviewing rounds.
- [Paper III] Hossein Nourollahi Hokmabad has co-authored the paper and provided data-driven non-intrusive load identification tool using machine learning techniques. He has also contributed to the paper writing and answering questions or concerns of reviewers during the peer-review process.
- [Paper IV] Hossein Nourollahi Hokmabad was a main author of the paper. He has developed a high-level control algorithm for phase balancing in the residential three-phase dwellings using single cell energy router topology. He has provided a literature analysis, calculation processes, and writing. He was responsible for submission and contact with editors during peer-reviewing round. The outcomes have been presented in IEEE Applied Power Electronics Conference and Exposition (APEC 2025), Atlanta, USA.
- [Paper V] Hossein Nourollahi Hokmabad was a main author of the paper. He has developed a forecast-driven scenario-based stochastic optimization framework for home EMS. Hossein Nourollahi Hokmabad, prepared the paper, the text, and developed the code and mathematical analysis. He presented the paper at the 19th IEEE International Conference on Control & Automation (ICCA 2025), Tallinn, Estonia.
- [Paper VI] Hossein Nourollahi Hokmabad was a main author of the paper. In this paper, he compared the economic and technical performance of various near zero energy building technologies such as pure direct current-based house, conventional houses, and hybrid alternatives and direct current-based topologies. Hossein Nourollahi Hokmabad, prepared the paper, the text. He presented the paper at the IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2024), Gdansk, Poland.
- [Paper VII] Hossein Nourollahi Hokmabad was a main author of the paper. He has developed a day-ahead solar forecasting method based on XGBoost and LSTM models. He has provided a literature analysis, model development, assessment, performance comparison, and writing. Hossein Nourollahi Hokmabad, prepared the paper, the text. He presented the paper at the IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE 2023), Dubrovnik, Croatia.

Abbreviations

DRES	Distributed Renewable Energy System
PV	Photovoltaic
EMS	Energy Management System
RES	Renewable Energy Source
V2X	Vehicle-to-X
EnergyR	Energy Router
LINES	Low-INertia System
nZEB	nearly Zero-Energy Building
ESS	Energy Storage System
MILP	Mixed-Integer Linear Programming
AI	Artificial Intelligence
DL	Deep Learning
ML	Machine Learning
LSTM	Long Short-Term Memory
SoH	State of Health
EV	Electric Vehicle
XGBoost	eXtreme Gradient Boosting
IoT	Internet of Things
GA	Genetic Algorithm
PSO	Particle Swarm Optimization
MPC	Model Predictive Control
RL	Reinforcement Learning
LP	Linear Programming
MINLP	Mixed-Integer Nonlinear Programming
QP	Quadratic Programming
AC	Alternating Current
DC	Direct Current
HPPU	High-Performance Processing Unit
BES	Battery Energy Storage
NWP	Numerical Weather Prediction
API	Application Programming Interface
HTTP	Hypertext Transfer Protocol
MQTT	Message Queuing Telemetry Transport
SAED	Self-Attention Encoder Decoder
WMO	World Meteorological Organization
BiLSTM	Bidirectional Long Short-Term Memory
SGDR	Stochastic Gradient Descent Regressor
nRMSE	Normalized Root Mean Square Error
nMAE	Normalized Mean Absolute Error
sMAPE	Symmetric Mean Absolute Percentage Error

MAPE	Mean Absolute Percentage Error
ARM	Auto Regressive Model
ARIMA	Auto Regressive Integrated Moving Average
SARIMA	Seasonal ARIMA
DSO	Distribution System Operator
NN	Neural Network
SoC	State of Charge
SoP	State of Power
DP	Dynamic Programming
CDF	Cumulative Distribution Function
SC-SP	Single Cell-Single Phase
TC-TP	Three Cell-Three Phase
EoL	End of Life
DT	Digital Twin

Symbols

I	Solar cell output current
V'	Output voltage
I_{ph}	Generated current by solar irradiance
I_d	Diode current
I_{pc}	Current of the parallel resistor
R_{pc}	Parallel resistor
R_{sc}	Series output resistor
I_o	Diode reverse saturation current
η	Ideality factor
q	Charge of the electron
k	Boltzmann's constant
T_c	Solar cell temperature in Kelvin (k)
T_a	Ambient temperature
T_{NOC}	Panel's temperature in C°
SI	Solar irradiance
N_s	Number of the cells in series in each single solar module
n	Number of forecasting data samples
y	Actual values
y'	Forecasted values
y_{max}	Maximum amount of the target value
SS_{res}	Sum of squares residuals
TSS	Sum of total squares
\bar{y}	Means of actual PV output power values
\bar{y}'	Mean of forecasted PV output power values
Gr_c	Utility usage cost
Es_c	Cost of energy storage fading
F_p	Amount of profit
n	Total optimization iterations steps
j	Iteration
\mathcal{F}	Vector of optimization variables
\mathcal{X}	Total optimization factors
G	Utility grid
ES	Energy storage
PV	Solar power
L	Demand or load
$P_{x/y}$	Power flow from node x to node y
ℓ	Total dimension of the optimization vector
\vec{E}_g	Net amount of electricity purchased from the main grid in each iteration
U_{ToU}	Vector of electricity tariff in each corresponding iteration

C_{fx}	Network service charge
E_{Ex}	Total amount of energy exchange with the grid in each iteration
C_{Ex}	Fixed cost of energy exchange with the grid
U_{BESS}	Unit price of home battery installation
C_{nm}	Home battery capacity
ΔSoH	Amount of home battery capacity degradation
SoH'_{j-1}	Energy storage SoH in the previous iteration
E_f	Array of total amount of dropped energy to the grid in each Optimization iteration
$P_{PV,total}$	Total amount of power generated by PV
$P_{L,total}$	Total amount of demand in the corresponding time interval
G_P^T	Total probabilities for production
D_P^T	Total probabilities for demand
\mathcal{F}_i	Vector of optimization variables
X_i	Vector of optimization factors
\hat{P}_i^T	Probability of each scenario's
\acute{s}	Number of selected candidate scenarios
ToU_h^{imp}	Time-of-use tariffs for imported energy
ToU_h^{exp}	Time-of-use tariffs for exported energy
$P_{A \rightarrow B}^i [h]$	Power flow from point A to point B
$P_{pv \rightarrow gr}^i [h]$	Power flow from PV to grid
$P_{pv \rightarrow es}^i [h]$	Power flow from PV to energy storage
$P_{pv \rightarrow ld}^i [h]$	Power flow from PV to load
$P_{pv,max}$	The maximum allowable power output from the PV system
$P_{gr \rightarrow ld}^i [h]$	Power flow from grid to load
$P_{gr \rightarrow es}^i [h]$	Power flow from grid battery storage
$P_{es \rightarrow ld}^i [h]$	Power flow from energy storage to load
$P_{es \rightarrow gr}^i [h]$	Power flow from energy storage to grid
SoC_{min}	Minimum allowable battery SoC levels
SoC_{max}	Maximum allowable battery SoC levels
$charge_{es}^i [h]$	The charging state of the battery
$discharge_{es}^i [h]$	Discharging state of the battery
C_{cal}	Calendar capacity loss
C_{cycle}	Capacity degradation
C_{nm}	Nominal capacity of the fresh battery cell
$C_{cal,fade}$	Percentage of total calendric capacity loss
m	Month
T	Ambient temperature
$SoC_{1h,avg}$	Average of SoC levels during each hour
$C_{cal,1h}$	Percentage of the capacity loss
h	Total hours of energy storage operation and relaxing times

Q_i	Amount of processed charge during each recharge/discharge cycle
T_i	Cell temperature at the i th charge or discharge moment
T_{ref}	Reference temperature
e	Total number of recharge and discharge cycles
SoC_{dev}	Normalized standard deviation of SoC
Q_m	Final amount of processed charge
Q_{m-1}	Initial amount of processed charge
ΔQ_m	Absolute amount of charge
π_Y	Total annual cash flow in the system,
ξ_c	Initial investment
Ω_{AN}	Annual inflation
ϑ_u	Remaining investment amount for compensation
ξ_A	Annual operating cost of the system
φ_k	Utility bill in the month number k
Θ_{PV}	Solar system size
U_*	Unit price
P_{max}	Selected energy router size
$\Omega_{*/size}$	Cost/size reduction rate
y	Running age of the setup in years
U_M	Annual maintenance cost ration
Ψ_*	The year in which the replacement occurred
Ω_{BESS}	Annual price reduction
$L_{1,t}$	Power consumption for phase 1 at time t
$L_{2,t}$	Power consumption for phase 2 at time t
$L_{3,t}$	Power consumption for phase 3 at time t
$L_{avg,t}$	Average power consumption from all phases in time t
$L_{unbl,t}$	Maximum power deviation
L''_y	Each phase's demand
L''_{avg}	Average demand
ES_{t+1}	Home battery next state
ES_{max}	Maximum energy capacity of the home battery
PV'_t	Remaining generated solar energy

1 Introduction

This thesis is devoted to developing a residential energy management platform, which transfers residential homes to a readiness level suitable for active integration into the future intelligent power networks, considering both hardware and software requirements.

1.1 Background

The power grid infrastructure, originally developed over two centuries ago to support centralized energy generation, has not kept pace with the growing electricity demand, the decentralization of power production, and shift toward Distributed Renewable Energy Systems (DRESs) [1]. As a result, power networks in high-demand regions, such as the European Union, are approaching their capacity limits, facing significant challenges in maintaining reliability and stability [2]–[5]. To address these issues and to support a seamless, efficient, and resilient energy transition, urgent and strategic upgrades are essential across all levels of the power system, from high-voltage transmission to low-voltage distribution [6], [7].

Modern technologies, including Electric Vehicles (EVs), data centers, digitalized urban infrastructure, automated manufacturing, and advanced agriculture, are highly dependent on electricity, driving a substantial and ongoing increase in power demand. Among these, buildings account for approximately 40% of total energy consumption, encompassing both electricity and heating [8]–[10]. The significant contribution of buildings to energy consumption establishes them as pivotal elements in power grid dynamics, prompting policymakers to implement stringent regulations on building energy efficiency and renewable energy generation capabilities. For instance, European regulations mandate that new buildings incorporate on-site renewable energy generation [11], with 95% of these structures employing solar Photo Voltaic (PV) systems [12]. As a result, buildings are transitioning from passive energy consumers to active participants in electricity production and reshaping the future grids.

On-site renewable energy generation, such as solar PV systems, enables local consumption and reduces reliance on long-distance transmission. However, its impact on the power grid is complex due to the intermittent and synchronized nature of solar production in near regions, coupled with misaligned electricity generation and consumption patterns. Consequently, buildings now exert a more significant strain on the electricity grid than previously observed [13]. A clear example is the “duck curve,” phenomenon in the highly penetrated DRESs grids, where neighbourhoods generate excess solar energy during sunny hours, lowering grid demand, but then need significant power when sunlight fades. Fig. 1.1 illustrates California's duck curve effect from 2015 to 2022. With the increased deployment of rooftop PV units, the midday demand valley deepens, approaching negative values, indicating that on-site distributed generation is exceeding local demand. This synchronized behavior challenges the maintenance of network voltage within acceptable limits and undermines grid reliability and resilience [14], [15]. Therefore, integrating additional Renewable Energy Sources (RESs) without effective coordination and smart infrastructure is no longer a simple plug-and-play solution [16]–[18].

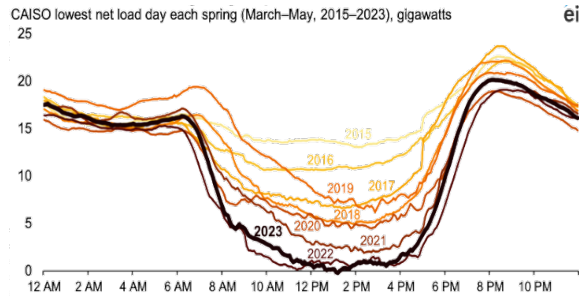


Figure 1.1 California’s duck curve trend with higher penetration of distributed renewable energy resources. CAISO lowest net load day each spring (March–May, 2015–2023) [19].

1.2 Motivation of the Thesis

As energy networks become more complex and evolve toward intelligent grids, buildings must also advance to be able to integrate with this ecosystem. Also, buildings with on-site renewable energy generation units and potentially equipped with battery storage technology require an intelligent Energy Management System (EMS) to coordinate and control energy generation, storage, and consumption.

Intelligent EMS aim to maximize energy self-sufficiency, minimize electricity costs, and ensure uninterrupted, green power delivery. Furthermore, these systems unlock new opportunities through emerging business models, services, asset sharing, and energy markets [20]–[22]. For example, well-equipped buildings can offer grid flexibility services [23], [24], form local energy communities [25], or operate as part of microgrids [26]. They can rent out their storage capacity [27], [28], provide EV charging infrastructure, or participate in Vehicle-to-X (V2X) interactions [29].

In response to these emerging needs, this thesis proposes the development of a platform designed to enhance building energy performance and facilitate effective integration of RESs and energy storages (home batteries) into buildings and low-voltage distribution networks. The work addresses both technical and economic challenges, offering a practical, and scalable solution. Moreover, it establishes a foundation for buildings to operate within smart grid ecosystems and future intelligent power networks.

This thesis was conducted as part of one of the research directions of the Power Electronics Group and software science department at Tallinn University of Technology. The aim is to gather knowledge and develop an advanced EMS solution for residential buildings. The current work was supported by: PRG675: New Generation of High-Performance Power Electronic Converters Simultaneously Applicable for DC and AC Grids with Extended Functionalities, EAG234: New Concept of Energy Router (EnergyR) for Residential Application (SoftER), PRG1463: Modeling and control of Low-INertia Systems (LINES), and TK230: Centre of Excellence in Energy Efficiency.

1.3 Aims, Hypothesis, and Research Tasks

This PhD research develops and validates an integrated hardware–software solution for next-generation building energy systems incorporating solar PV generation and battery storage units. The core software component is an EMS designed to intelligently manage and optimize power flow within buildings based on user-defined priorities, such as minimizing electricity bills or maximizing self-sufficiency. The primary hardware component is a power electronics device, termed the EnergyR, which provides a unified

interface for integrating PV systems and home batteries into the building and residential dwellings. The platform addresses key technical challenges, including system uncertainties, accurate forecasting of energy production and consumption trends, battery storage degradation effects, and the economic complexities of such systems. Validation is conducted through testing on a dedicated test bench at the TalTech nearly Zero-Energy Building (nZEB).

Hypotheses:

1. Economic barriers to widespread adoption can be alleviated by carefully sizing infrastructure such as solar PV units, power electronic devices, and Energy Storage Systems (ESS) to achieve minimal payback periods.
2. With certain modifications, single-cell power electronic interfaces can effectively connect solar PV units and ESSs to three-phase networks.
3. Stochastic optimization techniques can outperform deterministic methods by better accounting for the uncertainties inherent in building energy systems.
4. In solar systems with limited historical data, hybrid forecasting models—combining physics-based insights with data-driven techniques—can offer better day-ahead prediction accuracy.

Research Tasks:

1. Review the existing solutions in the market and determine the challenges in front of wide adoption of such systems.
2. Research and design a software tool for calculating the optimum infrastructure size to achieve lowest payback time based on buildings energy consumption profiles and solar power generation potentials.
3. Design an energy flow optimization algorithm and investigate various optimization methods, considering inherent uncertainties in the system.
4. Research and develop a day ahead solar power generation forecasting model considering residential roof-top setups limitations.
5. Build and prepare an experimental test setup for long term run test of the hardware and software in an online operation.

1.4 Research Methods

This PhD research employs mathematical analysis, simulation models, and experimental verification to develop and validate the proposed building energy system. A mathematical model of the system is formulated, incorporating a cost function that accounts for home battery capacity degradation, energy conversion losses, economic factors, and other relevant parameters, aligned with defined system objectives. The deterministic optimization problem is structured as a Mixed-Integer Linear Programming (MILP) model in the Pyomo environment and solved using the GLPK solver, implemented in Python. The stochastic optimization problem is also formulated as a MILP approach based on statistical features of historical data.

The optimization operates on a 24-hour day-ahead window, integrating a solar power generation forecasting unit. This forecasting model combines physics-based modeling with Artificial Intelligence (AI) techniques, including Deep Learning (DL) and regression methods. For load profile forecasting, two Machine Learning (ML) models are developed: Long Short-Term Memory (LSTM) networks and eXtreme Gradient Boosting (XGBoost) algorithms.

For hardware development, multiple tools and platforms are utilized. Altium Designer supports electronics hardware design, Raspberry Pi serves as an Internet of Things (IoT) server and runs optimization algorithms, and ESP32 facilitates home battery interfacing. MATLAB/Simulink is used for system modeling and simulation, while PVlib and Pybamm packages enable PV system and home battery modeling in Python, respectively. The imperial battery State of Health (SoH) estimation model is incorporated into the system. Experimental validation of theoretical results is conducted at TalTech's nZEB, equipped with an ESS, controllable loads, a solar PV system, power electronics infrastructure, and an EMS unit. The Power Electronics Research Laboratory at TalTech provides advanced facilities, including digital oscilloscopes, function generators, power quality and efficiency analyzers, microprocessor development tools, and PCB prototyping and assembly tools, supporting initial hardware development, home battery assembly, configuration, testing, and verification.

1.5 Contributions and Disseminations

The results of the research are presented via scientific publications, conferences, symposiums, doctoral schools, and presentations. During PhD studies the author contributed to 17 publications. Among them, five papers were published in peer-reviewed international journals. The remaining papers were reported at international IEEE conferences. The dissertation is based on nine main scientific publications, including three journals and six conference papers presented at IEEE international conferences.

Scientific novelties:

- Development of a novel day-ahead solar generation forecasting method for solar systems that addresses cold start issues in machine learning-based models.
- Justification of a novel single-cell three-phase topology, which reduces capital expenditure and improves the economic feasibility of home EMS systems.
- Development of an optimal infrastructure sizing algorithm that accounts for battery storage capacity fading.
- Proposed a scenario-based stochastic programming framework to optimize uncertain power flows within building electrical systems.
- Defined key functionalities and technical specifications required to prepare buildings for seamless integration into future intelligent energy networks.

Practical contributions:

- Experimentally tested the operation of single-phase power converters connected to a three-phase grid network and practically confirmed the applicability and benefits of the proposed solution.
- Built an experimental setup for running and testing EMS algorithms, including battery storage, a load simulator, and an EMS computational unit.
- Designed, developed, and programmed GUI on a 10-inch touch display, running on an Arm STM32H7 using the FreeRTOS operating system.
- Designed and developed a Wi-Fi-CAN interface for battery storage condition monitoring, data acquisition, and online control commands exchange.

1.6 Experimental Setup and Instruments

The main experimental tests have been performed in the nZEB house located on the campus of Tallinn University of Technology. Fig. 1.2a. shows the building outside view equipped with PV solar panels. Fig. 1.2b. shows the building inside view and installed infrastructure for EMS operation, and renewable energy source, and battery interfacing. Fig. 1.2c. shows a load simulator and various loads used during system tests. Fig 1.2 d. shows the home battery assembly and configuration stages which have been assembled and configured in the power electronics laboratory of Tallinn University of Technology. The oscilloscope Tektronix MDO4034B-3 with special probes Tektronix P5205A and Tektronix TCP0030A are used for voltage and current measurements, respectively. And the solar simulator equipment Chroma 62150H-1000S and battery simulator are used during battery operational verification tests.

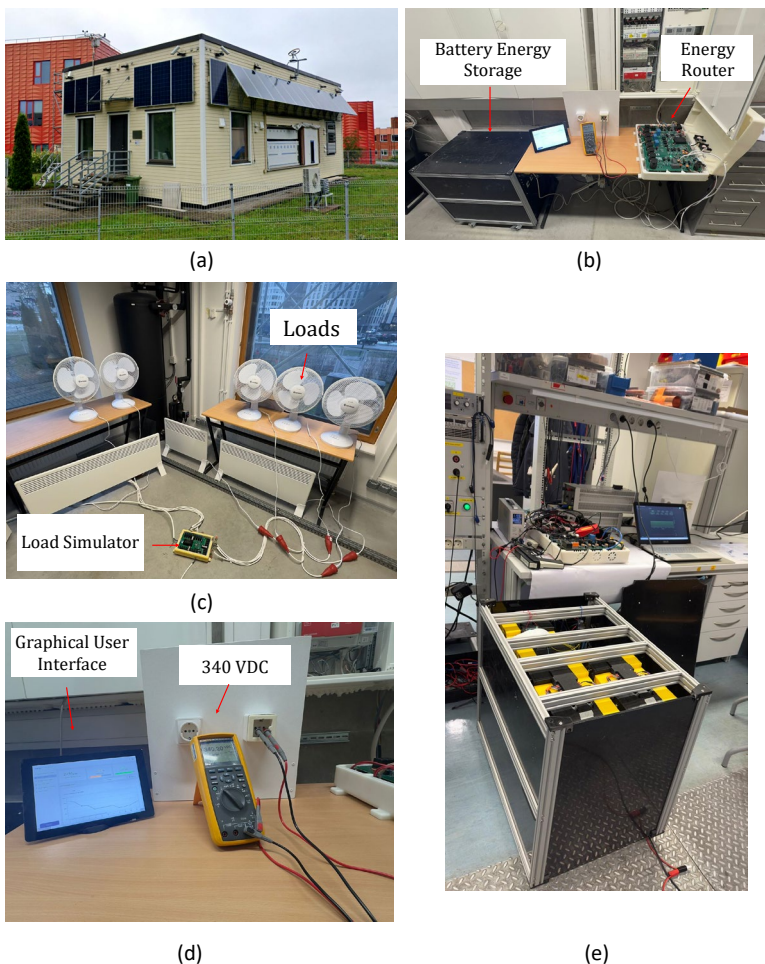


Figure 1.2 Experimental setup.

1.7 Thesis Outline

Chapter 2 introduces the structure of the developed hardware and software platform for the building EMS and describes the components in detail.

Chapter 3 details the developed tools for solar power generation and consumption forecasting, presents experimental results, and discusses the challenges in achieving accurate and stable outcomes.

Chapter 4 explains home EMS principles and compares the performance of ideal, deterministic, and stochastic optimization methods.

Chapter 5 presents the battery storage SoH optimization model, explains the optimal size calculation methodology, shows results based on the Estonian case study, and outlines strategies to reduce capital expenditure. Also, presents and discusses the economical and experimental justifications for a single-cell three phase-Energy Router for residential applications.

Chapter 6 presents the experimental results from EMS algorithm testing using the test bench and provides insights into the encountered challenges. Also, discuss the proposed hierarchical method for digital twinning of building energy management systems.

Chapter 7 presents a conclusion, learned studies related to building EMSs.

2 Proposed Hardware and Software Platform for Home EMS

This chapter aims to briefly review the existing commercial products and state-of-the-art solutions in academia and define the scientific and technical niches for the hardware and software components.

2.1 State-of-the-Art

Home EMS can be categorized into various groups based on their capabilities, objectives, and the employed software/hardware technologies. Table 2.1 summarizes different commercialized and proposed academic projects for home EMS solutions.

Table 2.1 Home EMS solutions comparison.

Reference	AC/DC	Optimization Method*	Forecasting	Hardware	Software	Load Scheduling	Grid Services	Battery Longevity	User Interface	Data Server	Cloud Service	Commercialized
[31]Tesla	AC	AI	✓	✓	✓	✗	✓	✗	✓	✓	✓	✓
Enphase	AC	RB+AI	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓
Huawei	AC	?	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Sonnen	AC	?	✗	✓	✓	✗	✗	✗	✓	✓	✓	✓
Schneider	AC	RB	✗	✓	✓	✓	✗	✗	✓	✓	✓	✓
[30]	AC	GA	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[31]	AC	PSO	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
[32]	AC	MPC	✓	✗	✗	✗	✗	✗	✗	✗	✗	✗
[33]	AC+DC	Ds-MPC	✓	✓	✗	✗	✗	✗	✗	✗	✗	
[34]	AC	DP	✓	✓	✗	✓	✗	✗	✓	-	-	✗
[35]	AC	RL	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
[36]	AC	FL	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
[37]	AC	MILP	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
[38]	AC	MINLP	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
[39]	AC	SP-PSO	✓	✗	✗	✗	✓	✓	✗	✗	✗	✗
[40]	AC	MADC	✗	✗	✓	✓	✗	✗	✗	✗	✗	✗
[41]	AC	ML	✓	✗	✓	✗	✗	✗	✓	✓	✗	✗
Proposed	AC+DC	MILP/SP	✓	✓	✗	✗	✗	✓	✓	✓	✓	✗

*RB: Rule-Based, GA: Genetic Algorithm, PSO: Particle Swarm Optimization, MPC: Model Predictive control, Ds-MPC: Distributed MPC, DP: Dynamic Programming, RL: Reinforcement Learning, FL: Fuzzy Logic, LP: Linear Programming, MILP: Mixed Integer Linear Programming, MINLP: Mixed Integer None-linear Programming, SP: Stochastic Optimization, MADDC: Multi Agent Decentralized Control, ML: Machine learning, AC: Alternative Current, DC: Direct Current.

As observed, the technology is rapidly commercializing, and companies in the renewable energy and home appliance sectors are introducing advanced solutions with features such as personalized load scheduling, user-tailored energy optimization algorithms, maximum energy autonomy, and grid independent energy supply. Since commercial products must ensure reliability and scalability, the employed management strategies are predominantly rule-based or follow user-defined logic. In certain cases, such as solutions from Tesla and Huawei, companies advertise “smart” home EMS capabilities but often do not disclose the underlying control complexity or optimization methodologies. Fig. 2.1 shows some commercialized solutions currently available in the market.

The Enphase Energy System

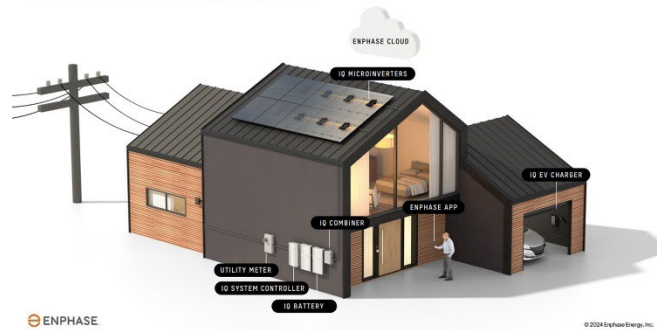


Figure 2.1 Some samples of commercialized solutions for home EMS from ENPHASE and LG ELECTRONICS companies.

In contrast with industry, academic research is largely focused on the development of advanced optimization and control methods for home EMS. These approaches can generally be grouped into four major methodological paradigms. The first includes heuristic and metaheuristic techniques such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization, which are widely used for their flexibility and effectiveness in solving non-convex problems despite lacking guarantees of global optimality.

The second paradigm is model-based control, where techniques like Model Predictive Control (MPC) are applied to optimize system performance over a prediction horizon using dynamic models of energy consumption and generation. The performance of predictive solutions greatly depends on how accurately they can predict the next states of the system. The third involves data-driven approaches, which leverage ML, DL specifically Reinforcement Learning (RL) to learn patterns from historical data or adaptively optimize decisions through interaction with the environment. The fourth paradigm

encompasses mathematical optimization methods, such as Linear Programming (LP), MILP, Mixed-Integer Nonlinear Programming (MINLP), and Quadratic Programming (QP).

Additionally, optimization strategies used in Home EMS can be further characterized as deterministic or stochastic, depending on whether uncertainty is explicitly considered in the formulation. From a system architecture perspective, control frameworks can be centralized, or decentralized, distributed, or even implemented as multi-agent systems, where multiple controllers coordinate actions based on partial or local information.

2.1.1 Research Niches

Based on the literature review, the main limitations are as follows: First, although numerous optimization algorithms have been proposed, most are evaluated solely in simulation environments with limited experimental validation. Second, the economic challenges associated with system deployment and operation, such as installation costs, energy market participation, and return on investment, are often overlooked. Third, the effects of home battery degradation and the development of degradation-aware EMSs have received limited attention.

Fourth, most of the existing research focuses predominantly on Alternative Current (AC) technologies, with minimal consideration of emerging Direct Current (DC)-based solutions. Given that DC technology is increasingly regarded as a key enabler for facilitating a smoother energy transition, DC-aware design, both at the software control level and in the supporting hardware architecture, is essential. Finally, few studies have attempted to develop integrated hardware and software solutions. Most research efforts focus on one aspect of the system. A holistic design approach that integrates both domains is critical to improving overall system performance, efficiency, scalability, and enabling more effective coordination and synergy among components.

2.1.2 Technical Niches

From a technical perspective, the main limitations are related to enhancing EMS performance and reducing the capital costs associated with system installation. One key area for improvement is the integration of accurate personalized, user-specific forecasting models capable of predicting short-term on-site power generation and demand. Such models can significantly enhance the EMS's ability to make future-aware decisions, thereby reducing operational uncertainty.

On the hardware side, innovations that lower system costs, improve energy efficiency, and minimize power conversion losses are particularly valuable. Furthermore, considering the growing relevance of DC technologies in future energy systems, it is essential to develop solutions that are compatible with DC-based technologies. Finally, developing real test sites and providing infrastructure for running experiments test could enhance the research and development quality.

2.2 Proposed Home EMS

Fig. 2.2 presents a holistic picture of the proposed home EMS ecosystem, comprising hardware components such as a High-Performance Processing Unit (HPPU), EnergyR, Battery Energy Storage (BES), solar photovoltaic (PV) systems, and Local database (Local DB). In addition, the architecture integrates cloud services for hosting software and AI functionalities and employs dedicated firmware to support edge computing capabilities at the device level.

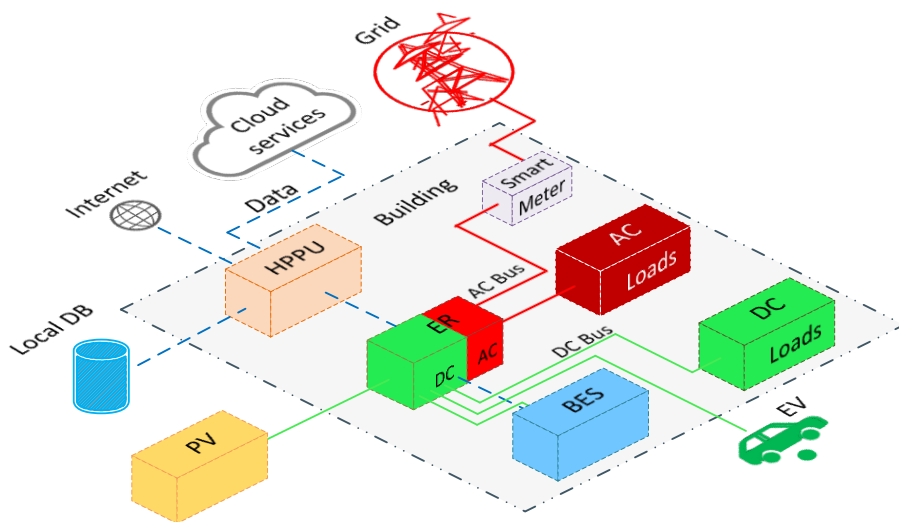


Figure 2.2 Proposed building DC-aware EMS ecosystem. EnergyR (ER).

The hardware solution and power electronics infrastructure are based on a DC-aware design philosophy, anticipating the growing adoption of this technology alongside traditional AC systems. Given the efficiency and performance advantages of DC solutions, the proposed EMS ecosystem is engineered, to operate in purely AC or hybrid AC/DC environments.

Fig. 2.3 illustrates the system architecture in greater detail. In this thesis, a Raspberry Pi 4 module is employed as HPPU, and the main processing unit. This device is responsible for handling computationally intensive tasks, such as executing real-time optimization algorithms. The HPPU functions as an edge device, enabling on-site power flow optimization and local data processing. However, the solar power generation and demand forecasting models are hosted in the cloud. These models receive Numerical Weather Prediction (NWP) data from external weather service providers through Application Programming Interfaces (APIs). The inference process is performed in the cloud platform, and only the resulting forecasts are transmitted to the edge device.

The HPPU also serves as the communication hub and Local DB, interfacing with other system components such as the energy router, home battery, and user interface dashboard. Communication is established over a Wi-Fi network using Hypertext Transfer Protocol (HTTP) requests and Message Queuing Telemetry Transport (MQTT) messaging protocol.

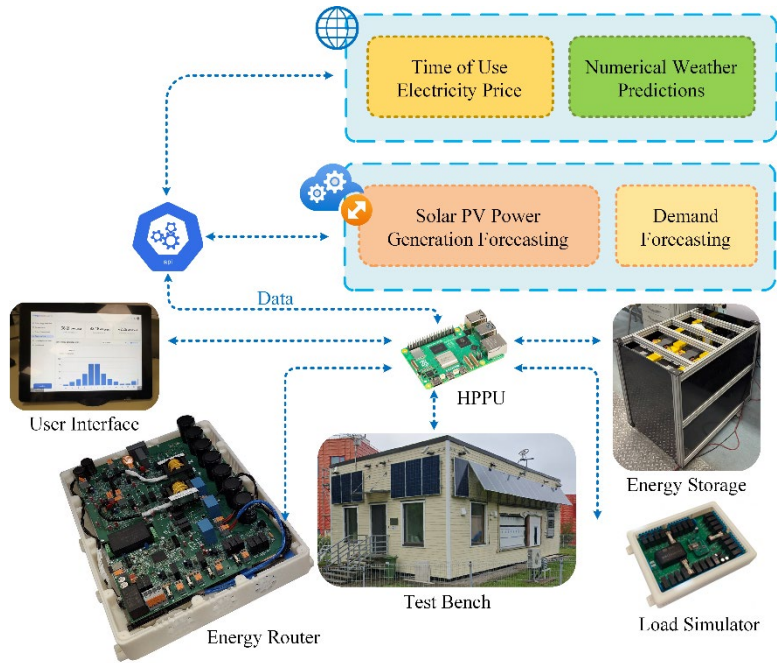
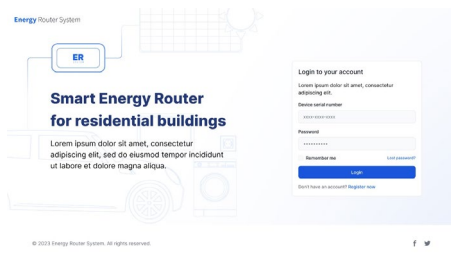


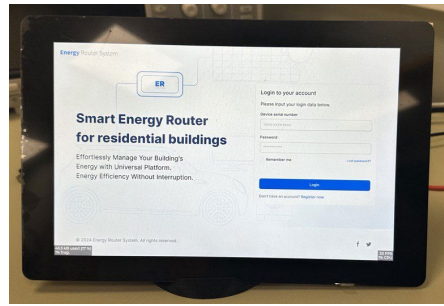
Figure 2.3 Proposed system components from data flow perspective.

For demonstration and monitoring, real-time system status, such as solar power generation, household consumption, electricity tariffs, economic performance, and home battery condition, is visualized through both a web-based dashboard and a local display. Users can interact with the system via these interfaces to adjust parameters and set optimization preferences. For example, they may choose to prioritize energy autonomy or focus on maximizing financial savings by exporting surplus energy to the power grid.

A local GUI displayed on a 10-inch LCD screen powered by an ARM-based STM32H7 microprocessor running a real time operating system called FreeRTOS. This local dashboard communicates with the HPPU over an internal Wi-Fi connection. Fig. 2.4.a shows the web-based user interface, and Fig. 2.4.b shows the local edge GUI. Fig. 2.5 shows the different pages of the designed GUI, and the navigation panel.



(a)



(b)

Figure 2.4 (a) Web based EMS user interface and dashboard, (b) Local GUI running on edge device for user interaction with a system.

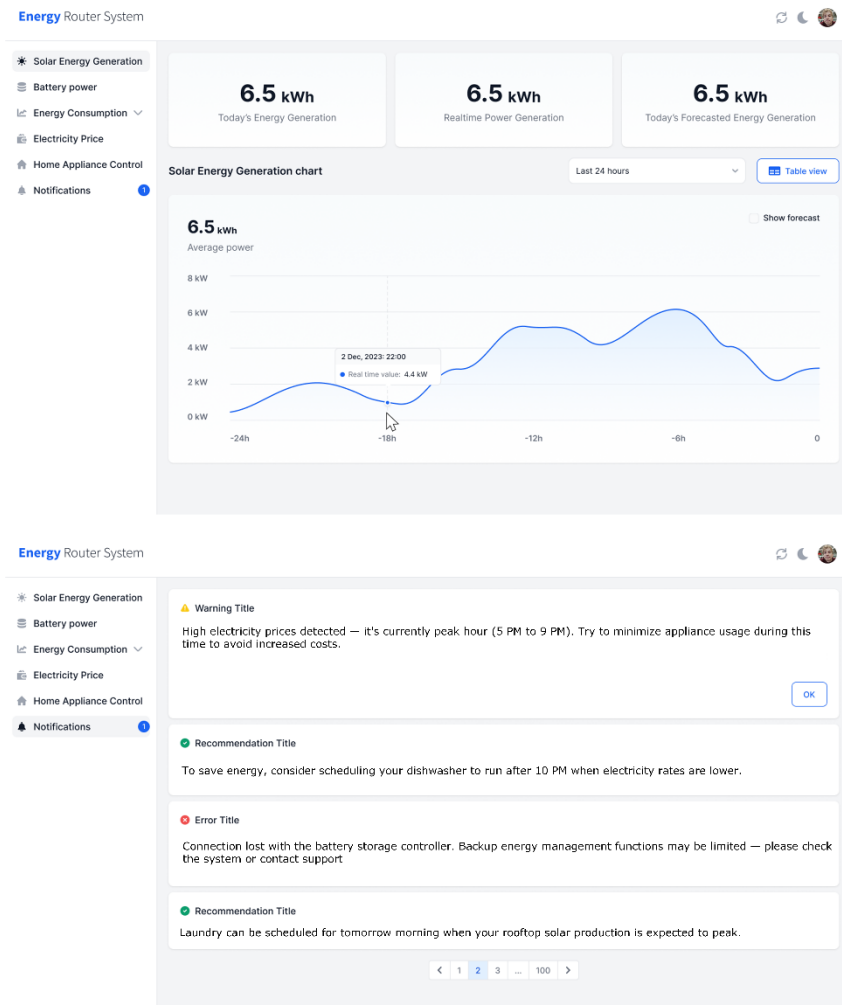


Figure 2.5 Designed user interface panel, navigation bar, and different sections for user interaction and system status reporting.

Following the identification of niches and projected future trends in Home EMSs, the proposed software platform has been structured with the HPPU serving as a gateway between the energy router, home battery, local DB and external internet and cloud services. This architecture enhances system security by preventing unauthorized access to private data, system statuses, and underlying hardware components without explicit permission from the HPPU. The software platform adopts a six-layer design structure, as illustrated in Fig. 2.6, with each layer responsible for a specific function.

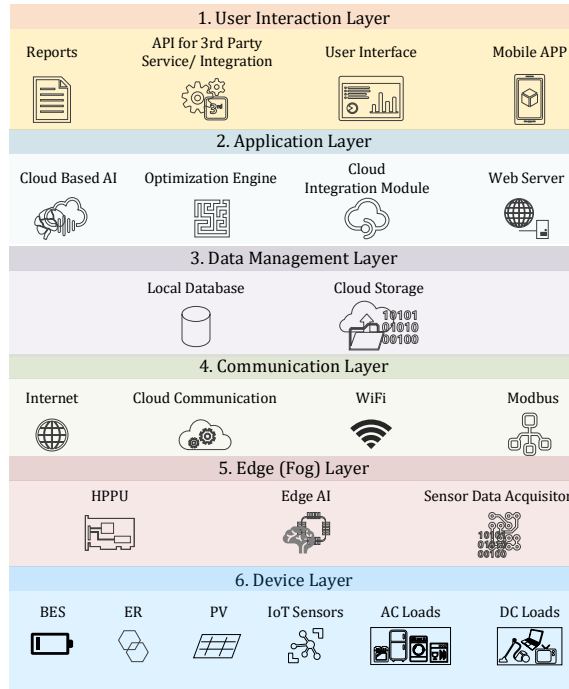


Figure 2.6 Proposed six-layer software architecture and each layers dedicated responsibilities.

This platform enables buildings equipped with PV and battery home to monitor, manage, and evaluate their internal power flows and overall performance. Moreover, it prepares buildings for seamless integration into future intelligent, renewable energy-driven power networks, enabling the houses to participate in emerging energy markets and local energy communities.

The backbone of the proposed EMS platform is a device called the energy router. This multi-port converter is capable of simultaneously delivering power to both AC and DC buses. Fig. 2.7 shows the EnergyR topology and its prototype. This device is a single-cell, three-phase system, meaning that instead of using a dedicated three-phase topology, a cost-effective single-cell topology is employed. The system includes controllable relays that allow the device to connect to the phase with the highest demand. This single-phase topology interacts with all phases, but not simultaneously. The technical and economic justification for the proposed solution will be discussed in chapter 5.

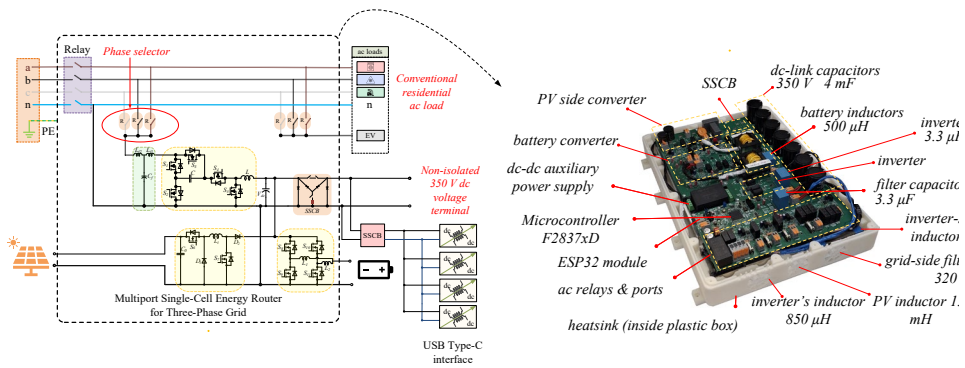


Figure 2.7 EnergyR internal topology and TRL-5 prototype.

During the experimental tests, human presence was not feasible; therefore, a load simulator was designed to generate realistic consumption patterns. These patterns were based on actual consumption data collected from a household with four residents in Estonia. To replicate the demand, various electrical loads with different power ratings were connected to the simulator. Fig. 2.8 shows the load simulator device, which consists of a controller board equipped with multiple mechanical relays for switching the loads on and off. Communication and control are handled via an ESP32-S3 module, which supports Wi-Fi connection and data exchange.



Figure 2.8 Demand simulation controller board.

For the home battery system, a custom-assembled and configured 11 kWh, 350 VDC lithium-ion battery bank is used. The battery bank consists of seven battery modules connected in series. Each module contains multiple battery cells arranged in a series-parallel configuration to meet the required voltage and current specifications. Fig. 2.9 illustrates the battery management system (BMS), battery wiring layout, and the home battery assembly process in the laboratory.

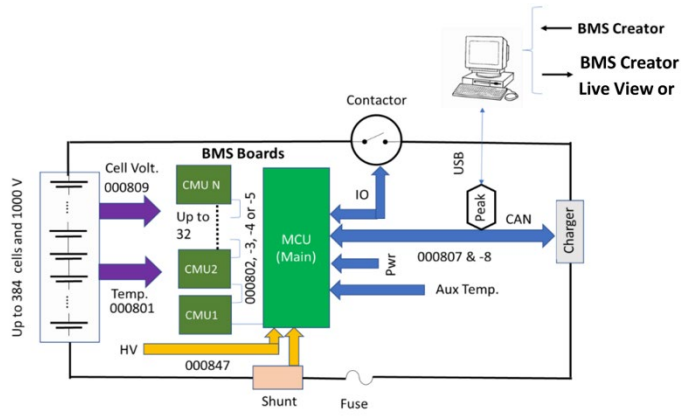


Figure 2.9 BMS configuration and connectivity, and home battery assembly and test stages in the laboratory.

2.3 Summary

In this chapter, the state-of-the-art academic and commercial solutions for home EMS are systematically reviewed, focusing on their architecture, optimization strategies, and functional features. A comparative analysis is presented, identifying their capabilities in load scheduling, battery management, forecasting, cloud integration, and user interaction. The chapter identifies key research and technical gaps. These include limited experimental validation, underexplored DC-based system architectures, inadequate economic analysis, and a lack of integrated hardware-software solutions. From a technical standpoint, challenges include improving system cost-efficiency, implementing personalized forecasting, and building real-world test infrastructures.

The chapter introduces a holistic home EMS ecosystem that integrates hardware and software components, designed for hybrid AC/DC operation to support emerging DC infrastructures. It is based on [43] and addresses research tasks 1 and 5. The main contribution is the proposal of a comprehensive DC-aware EMS solution, fully integrated into a hybrid nZEB.

3 Solar Power Generation and Electricity Consumption Forecasting

Accurate forecasting is essential for high-performance operation of home EMS. In building level EMS, two critical variables must be predicted: the renewable power generation and power demand. While data-driven methods such as AI based tools and ML models have demonstrated impressive accuracy in these tasks, they typically require extensive historical data for training. However, in most home EMS installations, such historical data is scarce or entirely unavailable. Yet, users or homeowners expect the EMS to operate properly immediately after deployment. This challenge requires innovative forecasting strategies that can operate effectively right from system’s launch and be able to offer reliable predictions without the need for long-term data accumulation.

3.1 Day Ahead Solar Power Generation Forecasting Algorithm

If we are not able to deliver an abundant amount of data for proper training of AI model, then the model will have an issue which is called cold start issue. In this case, the performance of model is poor since it was not able to acquire relevant knowledge from available data. One method to solve this issue is using hybrid techniques. One approach is combining AI based models, with physics-based ones. This synergy enables the system to deliver reasonable accuracy during the early data-scarce phase, while continuously enhancing its precision as real-world data is accumulated over time. Fig. 3.1 shows the proposed framework which combines physics-based and data driven approaches to reach better performance in various conditions.

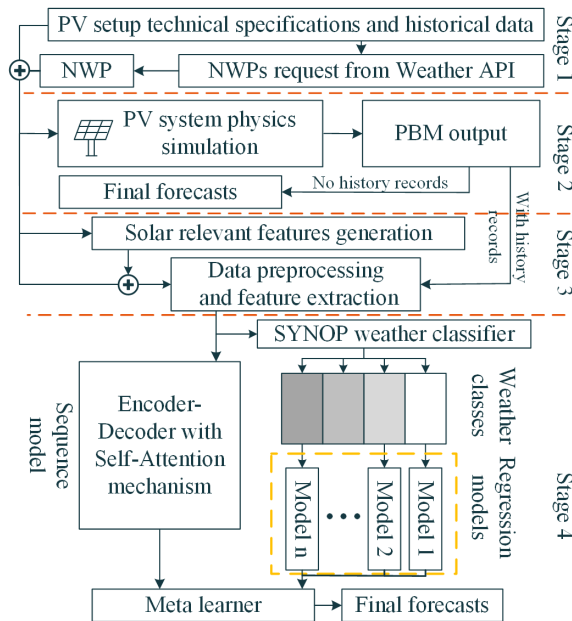


Figure 3.1 Proposed framework for day-ahead solar generation forecasting.

3.1.1 Physics-Based Model

Physics-Based Models (PBMs) use fundamental principles to explain system behaviours, relying on understanding underlying physical processes. The prevalent technique for solar cell simulation is the five parameters model, which characterizes the solar cell behaviour with a single diode equivalent circuit model. Additional components inherent to this model include an irradiance controlled current source, shunt, and series resistors [42]. The mathematical relation between cell current and voltage is given as:

$$I = I_{ph} - I_d - I_{pc} \quad (3.1)$$

$$I = I_{ph} - I_0 \left[\exp \left(\frac{q(V' + R_{sc}I)}{\eta \kappa T_c} \right) - 1 \right] - \frac{V' + R_{sc}I}{R_{pc}}, \quad (3.2)$$

where I is the solar cell output current, V' is the output voltage, I_{ph} is the generated current by solar irradiance, I_d is the diode current, I_{pc} is the current of the parallel resistor (R_{pc}), R_{sc} is the series output resistor, I_0 is diode reverse saturation current, η is the ideality factor, q is the charge of the electron ($1.602 \times 10^{-19} C$), κ is the Boltzmann's constant ($1.381 \times 10^{-23} \frac{J}{K}$), and T_c is solar cell temperature in Kelvin and can be calculated as [44]:

$$T_c = T_a + \frac{T_{NOC} - 20}{800} \times SI, \quad (3.3)$$

where, T_a is the ambient temperature (K), T_{NOC} is panel's temperature in C° in normal operation condition (extracted from panel datasheet), and SI refers to the solar irradiance ($W.m^2$). In PV panels, solar cells are connected in series/parallel formats. In series style, PV module's I/V relation can be calculated as:

$$I = I_{ph} - I_0 \left[\exp \left(\frac{q(V + N_s R_{sc} I)}{N_s \eta \kappa T_c} \right) - 1 \right] - \frac{V + N_s R_{sc} I}{N_s R_{ps}} \quad (3.4)$$

where $V = V' \cdot N_s$, and N_s denotes number of the cells in series in each single module. Fig. 3.2 shows the PV module layout connected in the series format.



Figure 3.2 PV module layout and equivalent circuit model of solar cells connected in the series format.

The output power of a PV system is not only influenced by the characteristics of the solar panels but also by the performance of the ancillary infrastructures, like solar inverters. Consequently, the power generated by a PV system at the maximum power point (MPP) for an individual panel can be approximated as [45]:

$$P_{pv} = \left[P_{pv}^{STC} \times \frac{SI}{1000} \right] \times [1 - \gamma \times (T_j - 25)], \quad (3.5)$$

$$T_j = T_{amb} + \frac{SI}{800} \times (N_{OCT} - 20) \quad (3.6)$$

where P_{pv} represents the panel's output power at MPP and P_{pv}^{STC} represents the output power in the standard test condition, γ is a temperature coefficient at MPP, T_j is the solar panel temperature ($^{\circ}C$), T_{amb} is ambient temperature ($^{\circ}C$), and N_{OCT} is a constant value.

3.1.2 ML Based Model

A major challenge for getting sufficient results from PBMs lies in their heavy reliance on accurate input data. As illustrated in Fig. 3.3, weather forecasts for the same location and time can vary considerably between different weather forecast service providers. This inconsistency severely limits the reliability of PBM outputs, whereas data-driven models are more adaptable to such variations, since they can learn and adopt themselves for these inconsistencies. In sequence data and time series trend forecasting, ML advancements, particularly DL, have significantly improved forecasting accuracy by effectively capturing complex, hidden patterns between input features and target outputs. In solar power generation forecasting, DL models excel at learning such dependencies, making them more robust than PBMs, when facing disturbances.

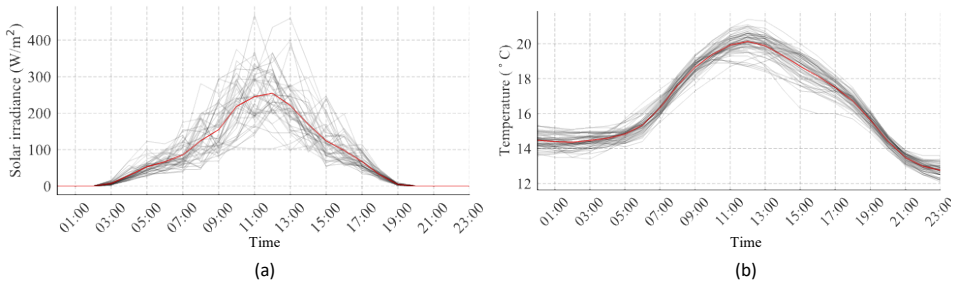


Figure 3.3 Solar radiation predictions for the same day and location, collected from different weather service providers (a). Predicted temperatures for the same day and location collected from different weather service providers (b).

The ML framework proposed in this research is illustrated in Fig. 3.4. The pipeline consists of two parallel branches: one based on classified regression models, and the other utilizing an encoder–decoder architecture with a self-attention mechanism. The outputs of both branches are then fused using a meta-learner model to produce the final forecast.

In this platform, the regression branch employs several LightGBM models, each optimized for specific weather conditions categorized using SYNOPTIC (SYNOP) weather codes. These codes, standardized by the World Meteorological Organization (WMO) and used globally by meteorological stations, classify weather into 100 distinct types, however not all categories are observed in every location. Using SYNOP codes for weather classification offers several advantages:

1. No need for additional clustering: Eliminates the computational cost and complexity of separate classification or clustering steps.
2. Universally recognized: Ensures consistency and interpretability across different regions and datasets.
3. Data-efficient: Particularly suitable for situations where historical data is scarce.
4. Reduced overfitting risk: Minimizes classification errors and overfitting compared to learned or dynamic clustering methods.

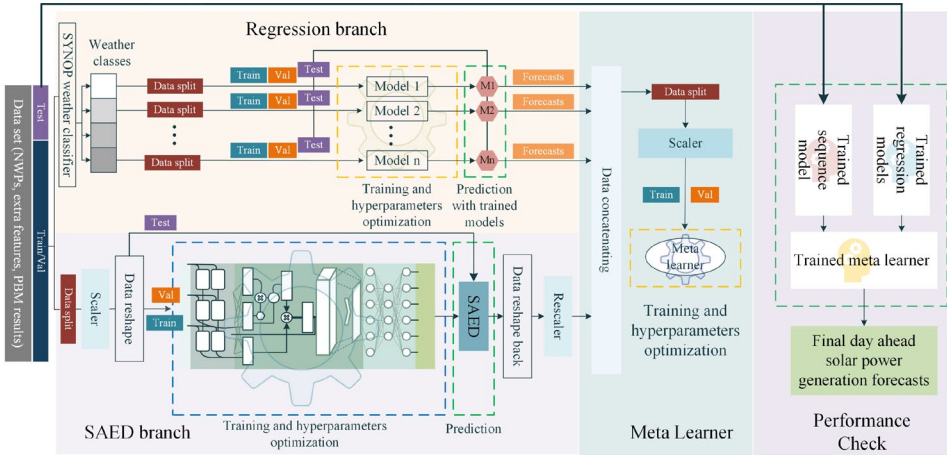


Figure 3.4 The detailed pipeline of the proposed hybrid ML model.

In the proposed ML pipeline, the sequence data analysis branch is composed of a Self-Attention Encoder-Decoder (SAED) network, as shown in Fig 3.5.

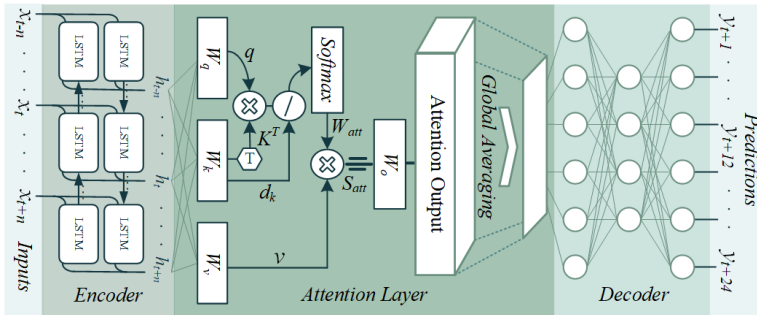


Figure 3.5 Proposed SAED based solar power forecasting model.

For the encoder layer, for encoding the input sequence data, a Bidirectional Long Short-Term Memory (BiLSTM) network is implemented. BiLSTM is capable of processing sequence data in both forward and backward directions, thus improving the model's insight into sequential data. To prepare the dataset for the BiLSTM layer, transforming it into a 3-D tensor shape is essential. This transformation was achieved using a sliding window technique, as illustrated in Fig. 3.6. In our approach, we utilized a look-back window of 360 records, corresponding to a continuous past 15 days period. Additionally, the sliding window is designed to move 24 steps each time, providing a new data segment for each day. The prediction horizon is set to 24 hours, aligning with the goal of day-ahead forecasting.

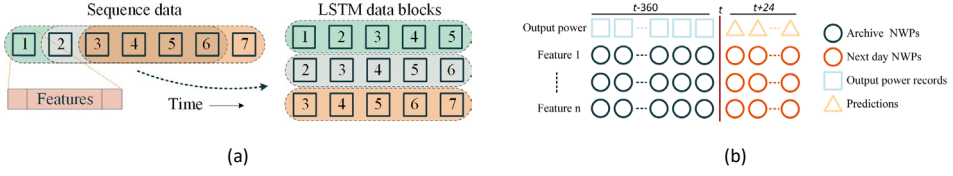


Figure 3.6 Sliding window technique and SAED network input data structure.

In this framework, a linear regression model has been trained using a Stochastic Gradient Descent Regressor (SGDR) [46] as the meta learner. This type of regression is adept at fitting by minimizing a regularized empirical loss through Stochastic Gradient Descent (SGD). The choice of SGDR is due to its online learning capability, which enables iterative model updates with new data inputs. This adaptability is crucial for accommodating changes resulting from environmental variations or alterations in system components. While there are other viable methods, such as weighted averaging, the use of SGD for training strikes offers a balance between performance and computational efficiency.

3.1.3 Numerical Results

Regression and forecasting model evaluations commonly utilize metrics like normalized Root Mean Square Error ($nRMSE$), normalized Mean Absolute Error ($nMAE$), coefficient of determination (R^2) score, Symmetric Mean Absolute Percentage Error ($sMAPE$), and Mean Absolute Percentage Error ($MAPE$). Notably, MAPE and sMAPE can pose challenges, especially when both the actual target and the forecasted value are zero. To mitigate this issue, MAPE and sMAPE in this study are calculated using a method recommended in [47]. The formulas for these metrics are detailed accordingly:

$$nRMSE = \frac{1}{y_{max}} \sqrt{\frac{1}{n} \sum_{h=1}^n (y_h - y'_h)^2} \quad 3.7$$

$$MAE = \frac{1}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{y_{max}} \quad 3.8$$

$$R^2 = 1 - \frac{SS_{res}}{TSS} \quad 3.9$$

$$MAPE = \frac{100}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{\bar{y}} \quad 3.10$$

$$sMAPE = \frac{100}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{\left(\frac{\bar{y} + \bar{y}'}{2}\right)} \quad 3.11$$

where n is number of forecasting data samples, y and y' are actual and forecasted values, respectively. y_{max} is the maximum amount of the target value in the corresponding test set. Also, $SS_{res} = \sum_{h=1}^n (y_h - y'_h)^2$ and $TSS = \sum_{h=1}^n (y_h - \bar{y})^2$ are sum of squares residuals and total sum of squares, respectively. \bar{y} and \bar{y}' represents the

mean of actual PV output power values, and the mean of forecasted PV output power values, respectively.

The performance of the proposed model was assessed in comparison to a spectrum of foundational models, including LSTM, recursive LSTM, BiLSTM, XGBoost, LightGBM, a basic encoder-decoder, averaging, and the Persistence model. The Persistence model employs a simple forecasting method, predicting that the PV power output for any given hour will replicate the output from the same hour on the previous day. Additionally, the average model generates forecasts by averaging the outputs from the two branches proposed in our model. Table I shows the comparison results and proposed models' superiority to its benchmarks.

Table 3.1 Comparison of forecast performances between sunrise and sunset periods.

Model	nMAE	MAPE %	sMAPE %	nRMSE	R^2 score
Proposed	0.043	22.62	21.62	0.089	0.898
Average	0.046	22.82	22.16	0.091	0.885
SAED	0.048	24.71	23.23	0.096	0.876
Classified LighGBM	0.050	25.54	24.63	0.101	0.864
LightGBM-LSTM	0.050	25.18	24.99	0.098	0.86
Recursive-LSTM	0.053	26.41	26.20	0.110	0.855
XGBoost	0.051	26.36	26.36	0.108	0.855
LightGBM	0.053	26.64	26.47	0.109	0.853
BiLSTM	0.056	27.88	26.93	0.114	0.829
Encoder-Decoder	0.053	26.77	26.41	0.111	0.853
LSTM	0.057	29.18	28.08	0.116	0.844
PBM	0.070	36.03	34.08	0.134	0.725
Persistence	0.084	43.13	43.13	0.174	0.584

The performance of the model across 15 consecutive days is presented in Fig. 3.7. It is apparent that the model's performance varies, underperforming on certain days, which highlights areas for further improvement. However, the overall efficacy for day-ahead predictions is satisfactory. During purely sunny days, the model's forecasting is near ideal, and it also performs well on mostly sunny days with minimal cloud coverage. Challenges arise in accurately predicting output during days with highly variable weather conditions. This issue is compounded by a decrease in the accuracy of NWP's under dynamic weather conditions. However, on rainy and cloudy days with full cloud coverage, where power generation is considerably lower, the model provides relatively close forecasts of the target values. Fig. 3.8 compares the residuals between actual measured values and the forecasts produced by the PBM and hybrid PBM-ML model. The comparison clearly demonstrates the superiority of the PBM-ML approach.

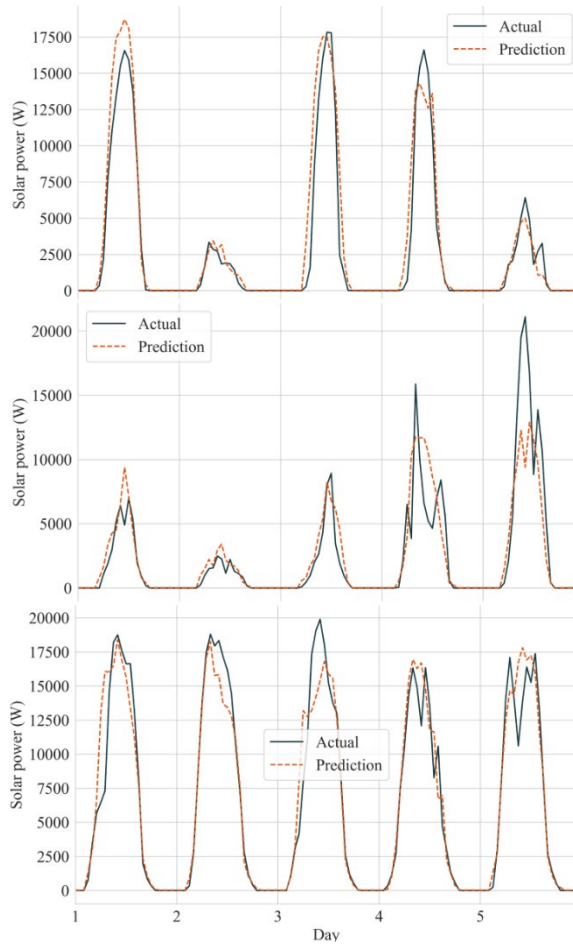


Figure 3.7 Proposed forecasting framework comparison with measured data.

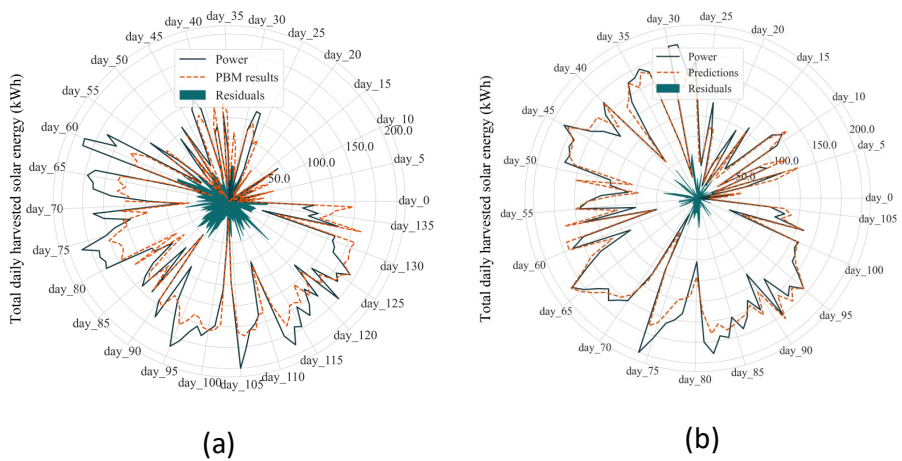


Figure 3.8 Comparison between performance of the hybrid ML-PBM model with PBM model. Hybrid Model (a), PBM model (b).

3.2 Day Ahead Electricity Consumption Forecasting Algorithm

Like solar power generation forecasting, having accurate insights into hourly electricity demand significantly enhances home EMS performance. However, demand forecasting poses greater challenges compared to solar forecasting, as it is closely tied to the behaviour and lifestyle patterns of home residents. Despite this complexity, the availability of electricity consumption data, which is usually recorded for billing purposes, alleviates the data scarcity issue. This makes the application of AI techniques more practical and effective. Additionally, traditional time-series analysis methods, such as Auto Regressive Models (ARM), Auto Regressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA), still are solid and reliable methods for achieving satisfactory results for demand forecasting.

In this thesis, the primary focus was placed on enhancing solar power generation forecasting, while a relatively straightforward approach was adopted for electricity demand forecasting. LSTM networks were selected for this purpose, given their proven effectiveness in handling sequential and time-series data, and their widespread use in related literature. Although this method provides a solid baseline, further research is required to achieve higher forecasting accuracy and robustness. In particular, transitioning from deterministic or single-point forecasts to probabilistic forecasting approaches could offer significant advantages by capturing the uncertainty and variability inherent in residential energy consumption patterns.

Fig. 3.9 presents the results of day-ahead, hourly-averaged demand forecasting using the LSTM network. While the predicted values do not precisely match the actual demand, the model demonstrates an ability to capture peak and valley patterns in the consumption profile and reasonably estimate their amplitudes. One significant challenge in improving model performance is access to high-quality data, which is often restricted due to privacy concerns. Since residential demand profiles are considered sensitive information, Distribution System Operators (DSOs) are generally prohibited from sharing them with third parties, as they may reveal personal habits and lifestyle information. However, emerging approaches such as federated learning offer promising solutions to this issue. Federated learning enables collaborative model training across multiple data sources without exchanging raw data and by only sharing Neural Network (NN) parameters, weights and bias values.

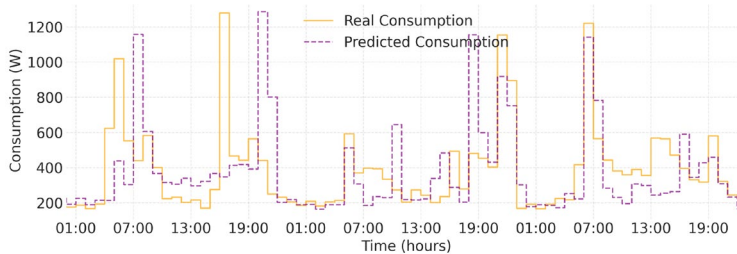


Figure 3.9 Hourly averaged electricity demand forecasting using LSTM network.

3.3 Summary

In this chapter, the developed approaches for forecasting solar power generation and household electricity consumption have been presented. Accurate forecasting is pivotal for efficient energy management, grid stability, and progress toward carbon neutrality. To address this need, the thesis introduces a hybrid day-ahead solar power forecasting model for behind-the-meter applications. This is achieved by combining PBMs with ML techniques. This approach accommodates both PV systems with extensive historical data and newly installed ones without prior records and mitigates cold-start issues.

For demand forecasting, a DL-based LSTM model is employed. This model demonstrates an ability to identify peaks and valleys in demand profiles. However, further research is required to improve demand forecasting accuracy. Novel approaches, such as federated learning, are needed to develop more precise and reliable models.

This chapter is based on paper II and paper VII and addresses research tasks 4 and hypothesis 4. The main contribution is development of a day-ahead solar power forecasting framework for residential small scale solar PV setups, preventing cold start issue when abundant historical data is not available for training complex ML algorithms. For this propose a physics-based solution was merged with data driven methods.

4 Developed Energy Management Method

Fig. 4.1 shows the general structure of the home EMS system. In the considered configuration, EMS is supervising and monitoring power flow inside the building by sending control commands to power electronic devices and battery management system. An effective energy management algorithm requires knowledge of the system's current state and predictions of its near-future conditions. This data helps the system to make informed, relevant, and optimal decisions.

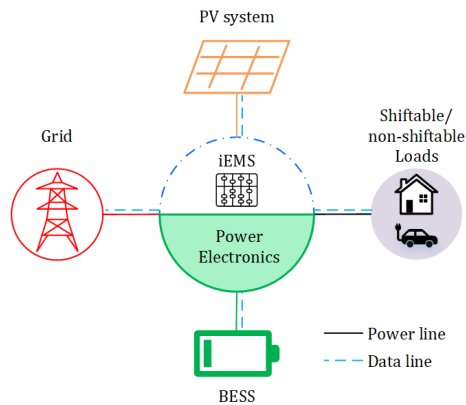


Figure 4.1 Home EMS system overview, considering integration of renewable energies, battery energy storage systems, and various loads.

A home EMS relies on specific parameters to determine the most suitable power usage strategy. These parameters can be categorized into two groups:

1. Near real-time parameters, including battery storage State of Charge (SoC), State of Power (SoP), and SoH, as well as real-time power demand, solar PV power generation, and power flow inside the building electricity network.
2. Future estimations, such as forecasted solar power generation, upcoming electricity tariffs, and predicted power demand for the next hours.

After collecting the necessary data, an appropriate cost function must be defined, considering user preferences, system priorities, and design constraints. In parallel, the building's electrical network needs to be accurately modelled using a set of mathematical equations that represent its operational characteristics. This modeling provides the foundation for the optimization algorithm, which aims to converge toward an optimal solution. The resulting strategy is tailored to meet specific objectives, such as minimizing energy costs or maximizing building energy autonomy. It should be noted that, in addition to any high-level, low-resolution EMS algorithm (which typically runs every few seconds or minutes), a low-level, high-resolution (micro/millisecond) control algorithm should be integrated into the power electronics devices, to ensure system correct and safe functioning. For instance, maintaining voltage, and current within an acceptable range, regardless of uncertainties in the system, and providing safety requirements.

In this thesis, in addition to the essential design parameters and constraints used for modelling and cost function formulation of home EMS, battery degradation is explicitly incorporated into the calculations. As the home battery is the most expensive component of the system, accounting for its degradation is critical for two reasons:

1. To prevent the EMS from over utilizing the home battery in energy arbitrage applications, which could accelerate degradation.
2. To enable the development of an aging-aware EMS, which is capable of effectively managing second-life batteries in residential applications. More detailed information about battery degradation modeling and its impact on EMS performance is shared in the next chapter.

At its core, any EMS serves as a framework for minimizing a defined cost function. Depending on specific goals, stakeholders, and operational priorities, this cost function can be formulated as either a single-objective or a multi-objective optimization problem. In the context of home EMS applications, the problem formulation can generally be summarized as follows: the EMS should aim to maximize the self-consumption ratio or minimize the electricity bill, while preserving user comfort and convenience. Additional functionalities, such as participating in energy markets or offering ancillary services, are typically considered secondary objectives in residential EMS solutions.

In addition, the optimization problem must account for physical and operational constraints, including the dynamic behavior and limitations of system components such as home battery, loads, and power electronics. To achieve this, home EMS approaches can be categorized into three main groups: 1- Heuristic (rule) based EMS, 2-Deterministic EMS, and 3- Stochastic EMS.

4.1 Heuristic (Rule) Based Approach

The heuristic, or rule-based, approach is widely used in commercial solutions due to its computational simplicity and straightforward implementation. These methods indirectly specify system models by implementing sets of rules related to system operation. The associated rules can be as simple as: *charge the BES when generation is higher than demand* or as complex as: *define the BES charging and discharging schedule based on real-time market prices, forecasted demand, and generation profiles*. However, formulating these complex rules, requires extensive knowledge such as power generation, and demand forecasts. Thus, in most cases rules are defined without including exogenous information, such as forecasts, or energy-market prices. Algorithm I shows the sample of applied rule-based approach in the literature and this thesis. In the proposed algorithm, R represents the renewable energy generation, and D denotes the energy demand. The parameter E_{extra} refers to the excess energy generated by the renewable source that cannot be consumed or stored at a given moment. D_{rem} represents the residual energy demand that cannot be fulfilled by the available renewable generation and must be covered by other sources such as home battery or grid.

4.2 Deterministic approach

This approach is gaining popularity for home EMS applications, particularly due to its simplicity. In such deterministic methods, decisions are made under the assumption that all system parameters are precisely known and there is no uncertainty. Although this assumption might initially seem unrealistic and potentially detrimental to model performance, yet it can be applied for high level optimization where the lower-level high resolution control system handles the inherent uncertainties. The reduced complexity makes these methods highly suitable for implementation on resource-constrained edge hardware platforms.

Algorithm 1 Heuristic-Based EMS

```
1: if  $R \geq D$  then
2:   Meet the demand using renewable generation
3:    $E_{\text{extra}} \leftarrow R - D$ 
4:   if  $E_{\text{extra}} > 0$  then
5:     while  $\text{SOC} < \text{SOC}_{\text{max}}$  and  $E_{\text{extra}} > 0$  do
6:       Charge the battery
7:       Update the SOC
8:        $E_{\text{extra}} \leftarrow E_{\text{extra}} - \text{Energy charged}$ 
9:     end while
10:    if  $E_{\text{extra}} > 0$  then
11:      Export  $E_{\text{extra}}$  to the grid
12:    end if
13:  end if
14: else
15:   Meet the demand using renewable generation
16:    $D_{\text{rem}} \leftarrow D - R$ 
17:   if  $\text{SOC} > \text{SOC}_{\text{min}}$  then
18:     while  $\text{SOC} > \text{SOC}_{\text{min}}$  and  $D_{\text{rem}} > 0$  do
19:       Use battery to meet remaining demand
20:       Update SOC and  $D_{\text{rem}}$ 
21:     end while
22:   end if
23:   if  $D_{\text{rem}} > 0$  then
24:     Import  $D_{\text{rem}}$  from the grid
25:   end if
26: end if
```

Popular solutions within this category include MPC, which optimizes energy consumption by forecasting near-future demand and supply while respecting system constraints, and applies only the first control action at each step before re-optimizing in the next time step (receding horizon). LP is also widely adopted, as it formulates energy management tasks as linear optimization problems aimed at minimizing costs or maximizing operational efficiency. Additionally, Dynamic Programming (DP) offers a structured approach by decomposing complex energy management decisions into a series of simpler, sequential subproblems, enabling optimal scheduling of energy resources over time.

In this thesis, a deterministic approach was initially developed using a combination of rule-based and nonlinear LP framework, specifically employing a MINLP. The choice of MINLP over MILP stems from the need to incorporate battery degradation modeling within the optimization loop. The battery degradation related equations introduce nonlinear dependencies into the system and cannot be accurately captured using linear techniques. Fig. 4.2 shows the applied method's visual pipeline.

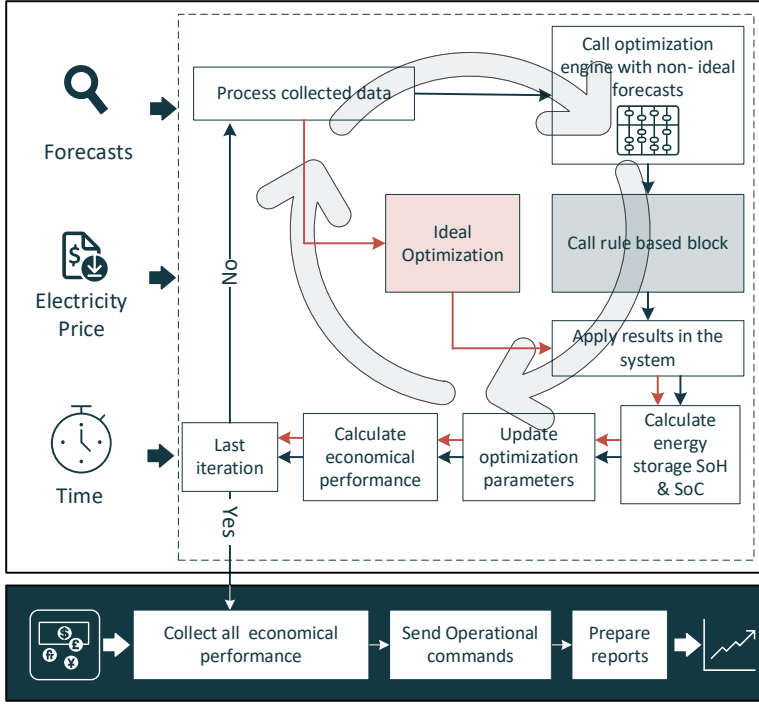


Figure 4.2 Deterministic Energy flow Optimization flowchart.

In Fig. 4.2, an ideal optimization block is employed to execute the same algorithm using perfect forecast data, revealing the theoretical upper bound of system performance as a benchmark for evaluating real-world outcomes. Since decision performance is inherently tied to forecast accuracy, comparing results under ideal and actual conditions offers valuable insights into the impact of forecasting inaccuracies. For instance, Monte Carlo sensitivity analysis can introduce controlled random noise into the forecast data, simulating various uncertainty scenarios. This enables an evaluation of the model's robustness and performance under diverse and imperfect conditions.

During system operation, the optimization problem should be continuously solved based on new conditions and subjections. As stated, in this thesis, maximizing profit considering home battery capacity fading, has been defined as the optimization problem's main objective. Thus, the system cost function is formulated as:

$$\min_x \mathcal{F}^T \cdot \mathcal{X} \rightarrow \min \sum_{j=1}^n Gr_{c,j} + Es_{c,j} - F_{p,j}, \quad (4.1)$$

where $Gr_c \in \mathbb{R}^n$ is the utility usage cost, $Es_c \in \mathbb{R}^n$ is the cost of energy storage fading, $F_p \in \mathbb{R}^n$ is the amount of profit system can earn by trading energy with the utility grid, n is the total optimization iterations steps, and j represents j^{th} iteration. In Eq. 4.1), $\mathcal{F} \in \mathbb{R}^\ell$ is the vector of optimization variables, and $\mathcal{X} \in \mathbb{R}^\ell$ represents the total optimization factors, $\vec{\mathcal{X}} = [\vec{P}_{PV/G}, \vec{P}_{PV/L}, \vec{P}_{PV/ES}, \vec{P}_{G/L}, \vec{P}_{G/ES}, \vec{P}_{ES/L}, \vec{P}_{ES/G}]^T$ where, suffix G represents utility grid, ES represents energy storage, PV represents solar power, and L represents demand or load. $P_{x/y} \in \mathbb{R}^n$ represents the amount of power flow from node x to node y , and ℓ represents the total dimension of the optimization vector. $Gr_{c,j}$ for residential sector is calculated as:

$$Gr_{c,j} = E_{g,j}U_{ToU,j} + C_{fx} + E_{Ex,j}C_{Ex}, \quad 4.2)$$

where $\vec{E}_g \in \mathbb{R}^n$ is the net amount of electricity purchased from the main grid in each iteration, $U_{ToU} \in \mathbb{R}^n$ is vector of electricity tariff in each corresponding iteration, $C_{fx} \in \mathbb{R}$ is the network service charge, $E_{Ex} \in \mathbb{R}^n$ is the total amount of energy exchange with the grid in each iteration. E_{Ex} is calculated accumulating total amount of imported and exported energy, and $C_{Ex} \in \mathbb{R}$ is the fixed cost of energy exchange with the grid.

Energy storage fading expenses ES_c , is calculated based on the capital cost of home battery installation and assuming that when the End of Life (EoL) point is met, the storage will not have any value for the user. Thus, the $ES_{c,j}$ is calculated as:

$$ES_{c,j} = U_{BESS}C_{nm} \left(\frac{100 - SoH'_{j-1}}{30} \right) \Delta SoH_j, \quad 4.3)$$

where U_{BESS} €/kWh is the unit price of home battery installation. Thus, $U_{BESS} C_{nm}$ is the total cost of adding energy storage to the dwelling. $\Delta SoH \in \mathbb{R}^n$ is the amount of degradation that occurs due to battery utilization in each optimization stage. SoH'_{j-1} is the energy storage SoH in the previous iteration. Consequently, the cost associated with utilizing battery rises as SoH decreases. Finally, $F_{P,j}$ is computed by:

$$F_{P,j} = (E_{f,j} \cdot U_{ToU,j}) \times \gamma, \quad 4.4)$$

where $E_f \in \mathbb{R}^n$ is the array of total amount of dropped energy to the grid in each optimization iteration, and $\gamma \in [0,1]$ is a Feed in Tariff (FiT) coefficient. Starting from the second half of 2022, the FiT is set to be 20% below the Time of Use (ToU) electricity purchase price.

Energy is the time integral of power, and under the assumption of constant time intervals and constant system voltage and current levels within each interval, energy parameters can be represented by power multiplied by time values $E = P \times t$. So, the optimization function can be reformulated and solved based on building grid's power values. For each infrastructure configuration the constraints will be different based on PV setup, home battery, and power electronics size. For instance, infrastructure sizes as 3 kWp, 5 kWh, and 5 kW respectively, the non-equality constraints for this setup will be as:

$$\forall j, 0 \leq P_{PV/G}[j], P_{PV/L}[j], P_{PV/ES}[j] \leq 3000, \quad 4.5)$$

$$\forall j, 0 \leq P_{ES/G}[j], P_{ES/L}[j], P_{G/ES}[j], P_{G/L}[j] \leq 5000, \quad 4.6)$$

$$\forall j, P_{G/L}[j] + P_{ES/L}[j] + P_{PV/L}[j] \leq 5000, \quad 4.7)$$

$$\forall j, P_{ES/G}[j] + P_{PV/G}[j] \leq 5000, \quad 4.8)$$

$$\forall j, P_{ES/G}[j] + P_{ES/L}[j] \leq 5000, \quad 4.9)$$

$$\forall j, P_{ES/L}[j] + P_{ES/G}[j] \leq (SoC[j] - SoC_{min}) \times C''_{nm} \times SoH[j], \quad 4.10)$$

$$\forall j, P_{PV/ES}[j] + P_{G/ES}[j] \leq (SoC_{max} - SoC[j]) \times C''_{nm} \times SoH[j], \quad 4.11)$$

$$\forall j, P_{PV/ES}[j] + P_{G/ES}[j] \leq 5000, \quad 4.12)$$

$$\forall j, P_{PV/G}[j] + P_{PV/L}[j] + P_{PV/ES}[j] \leq 3000, \quad 4.13)$$

$$\forall j, P_{G/ES}[j] + P_{G/L}[j] \leq 5000, \quad 4.14)$$

where, suffix G represents utility grid, ES represents energy storage, and L represents demand or load. $P_{x/y}[j]$ represents the amount of power flow from node x to node y in the house distribution system during j^{th} iteration. $C''_{nm} \times SoH[j]$ represents the amount of power, in watts, that the energy storage can consistently deliver for a duration of one hour.

The equality constraints are listed as follows:

$$\forall j, P_{PV/G}[j] + P_{PV/L}[j] + P_{PV/ES}[j] = P_{PV,total}[j], \quad 4.15)$$

$$\forall j, P_{PV/L}[j] + P_{ES/L}[j] + P_{G/L}[j] = P_{L,total}[j], \quad 4.16)$$

where $P_{PV,total}$ is the total amount of power generated by PV system and $P_{L,total}$ is the total amount of demand in the corresponding time interval. Nonlinear equations are introduced as:

$$\vec{P}_{G/ES} \cdot \vec{P}_{ES/G} = 0, \quad 4.17)$$

$$\vec{P}_{PV/ES} \cdot \vec{P}_{ES/G} = 0, \quad 4.18)$$

$$\vec{P}_{PV/ES} \cdot \vec{P}_{ES/L} = 0. \quad 4.19)$$

These limitations prevent the events in which the algorithm may decide to charge and discharge the energy storage simultaneously. Finally, $E_{f,j}$ and $E_{g,j}$, can be compiled as:

$$E_{f,j} = (P_{ES/G}[j] + P_{PV/G}[j]) \times t, \quad 4.20)$$

$$E_{g,j} = (P_{G/ES}[j] + P_{G/L}[j]) \times t. \quad 4.21)$$

The pre-defined nonlinear optimization problem was addressed using the Python Pyomo optimization toolbox and the 'fmincon' solver. This solver utilizes a gradient of the cost function to find the solution corresponding to objective function global minimum.

Including forecasting data can enhance the performance of EMSs. However, the designed cost function and the core optimization engine do not incorporate forecasts into their optimization principles. This limitation leads to decisions being made based on the current system state, restricting the benefits from future knowledge. To address this issue partially, a rule-based decision-making block has been added to the main optimization loop. The rule-based block utilizes day-ahead energy and load profile forecasts to generate charge and discharge signals for home battery when certain conditions are met. Thus, at the beginning of each day, total solar energy generation and total load demand for the next 24 hours are forecasted. Additionally, the minimum and maximum electricity prices and corresponding hours are extracted from the energy market data.

In moments with minimum energy prices, if one of the conditions below is met, the block sends a charge command to the home battery:

1. The forecasts for total generated energy should be less than the forecasts of total load demand (This confirms that the PV will not be able to handle all energy demand solely). In this stage, the battery can be charged in the amount to fill the existing gap.
2. $PR_{\max} - PR_{\min} \geq ES_{c,j}$, where PR_{\min} and PR_{\max} are the minimum and maximum prices of electricity on the following operation day, respectively (This

ensures that the cost of utilizing energy storage is less than the benefit from energy arbitrage). In this condition, the energy storage will be fully charged to maximize energy arbitrage.

In peak periods, if both conditions below are met, the block sends a discharge command to the home battery:

1. The stored energy in the battery plus the energy generation forecast should be higher than the load demand forecasts for the rest of the day (This ensures the energy storage has surplus energy for trading). In this condition, the energy storage is allowed to release surplus energy to the grid.
2. $PR_{\max} - PR_{\min} \geq Es_{c,j}$.

4.3 Probabilistic Approach

The probabilistic approach is designed to account for the inherent uncertainties in the system during operation and decision-making processes. By incorporating uncertainty, this method enables more informed and robust decisions, resulting in system behavior that aligns more closely with the planned or estimated performance. However, a major limitation of this approach is its computational intensity. For example, in scenario-based probabilistic optimization, the algorithm must evaluate multiple potential scenarios, each weighted by its probability of occurrence, and then determine an optimal solution that performs well across all scenarios, instead of optimizing for a single deterministic case.

In this thesis, a novel stochastic optimization approach based on nominated scenario generation considering both historical and forecasted data has been proposed. In this method, initially the demand and generation forecasts are derived from forecasting models which have been introduced in Chapter 3. As mentioned, forecasting tools generate 24 predictions for electricity generation and consumption (demand) for the next day. Let \bar{F}_h^G and \bar{F}_h^D represent the forecasted values for power generation and demand at hour h , respectively. Assuming the forecasting models provide acceptable accuracy; these values should serve as the best possible estimates of the system's uncertain input parameters for the next 24 hours. This assumption is the foundation of proposed stochastic optimization method. Fig. 4.3 shows the flowchart of a proposed method.

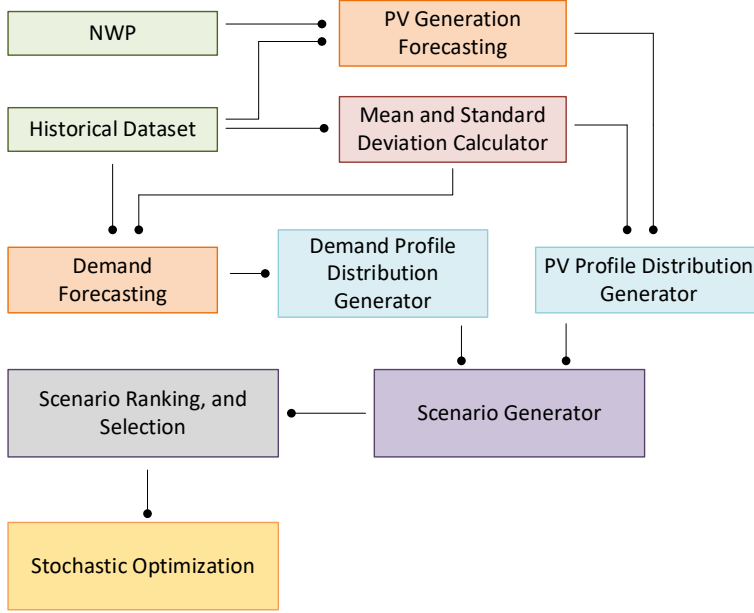


Figure 4.3 A flowchart of proposed stochastic optimization method based on forecast driven and historical data scenario generation.

Furthermore, based on available historical records and data distribution the mean and standard deviation (σ) are calculated for each hour of the day. After obtaining these values, the initial historical mean is replaced with the forecasted values (\bar{F}_h^G and \bar{F}_h^D). Assuming a Gaussian distribution, a normal distribution curve is then generated using the updated mean and the previously computed historical σ for each hour of the day. Since these curves represent physical quantities with finite values, they are constrained within the minimum and maximum possible ranges for electricity generation and demand.

Then, the covered range is divided into $\mathcal{R} = 100$ sections, and for each section, the probability of the actual measured value falling within that range is calculated. For instance, assuming the maximum power generation capacity of the solar PV system is 5 kWp, each section will have a resolution of 50 W. By limiting the number of sections to a fixed value, regardless of sizes of PV systems and building demand, the computational complexity remains consistent across all cases. By dividing the continuous range of possible values, we discretize and limit the potential subsequent values. However, the impact of this action is negligible in system performance.

After calculating these probabilities, a matrix of forecasting probabilities is constructed. Let $\mathcal{G}_{p\mathcal{F}}$ and $\mathcal{D}_{p\mathcal{F}}$ represent the matrices for the probabilities of potential values for electricity generation and demand, respectively, with dimensions $\mathcal{R} \times \mathcal{D}$, where \mathcal{D} is the forecasting horizon, which in this case equals 24. Similarly, a probability matrix is generated for each hour of the day using previously recorded historical data on solar PV power generation and demand. Let $\mathcal{G}_{p\mathcal{H}}$ and $\mathcal{D}_{p\mathcal{H}}$ represent the matrices for the probabilities of historical values for generations and demand, respectively, with dimensions $\mathcal{R} \times \mathcal{D}$.

The calculation of the $\mathcal{D}_{p\mathcal{H}}^{\mathcal{H}}$ matrix is straightforward. First, recorded demand values are clustered based on their respective hours of the day and then grouped according to their corresponding power range. Once all records are classified, the probabilities for each

hour and power range are computed. However, since solar PV power generation is highly dependent on weather conditions, classifying records based solely on temporal data would lead to inaccurate results, as solar irradiance in summer is not comparable to that in winter. To address this, the seasonality impact is eliminated based on a proposed solution.

To achieve this, for each day of the year and each hour of the day, the maximum possible solar irradiance values are calculated based on the sun position in the sky, and building’s latitude, and longitude. These values represent the theoretical maximum possible generation under clear sky conditions. The obtained solar irradiance value is then fed into a physics-based simulation of the building’s solar PV system to determine the potential maximum power generation. The minimum possible solar PV power generation value is derived from historical data by identifying the lowest recorded value for the same day and hour within a ± 30 -day window.

Using the obtained minimum and maximum ranges for each hour and day of the year, recorded solar PV power generation values can be categorized into a predefined number of classes. This is done by normalizing the range, determining class boundaries, and assigning each measurement to its corresponding class. In this approach, measurements are classified based on the percentage of solar PV power generation relative to the feasible range for each hour and day. This eliminates the seasonality factor from the data, allowing for a direct comparison of solar PV power generation probabilities between summer and winter without considering the absolute magnitude of the data. Fig. 4.4 illustrates the process described.

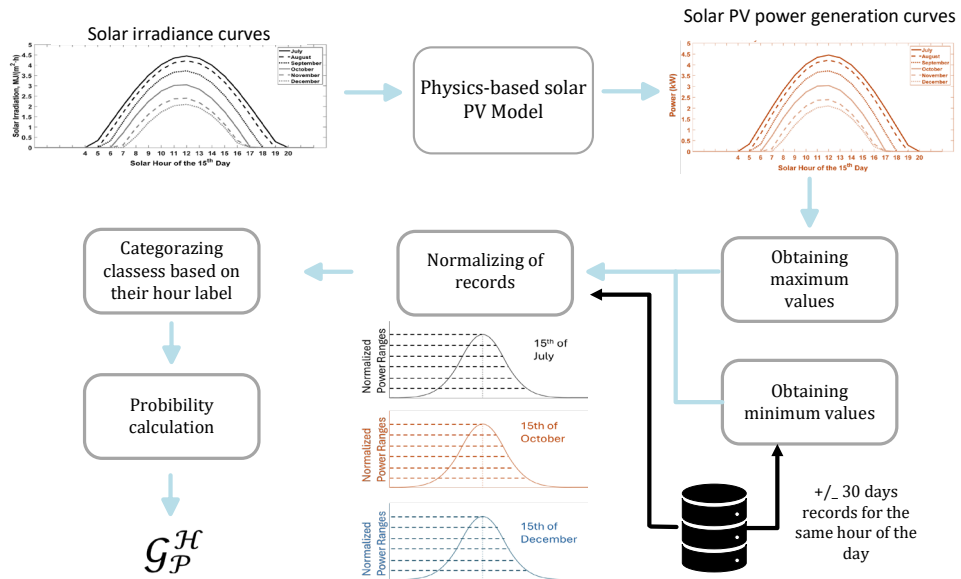


Figure 4.4 Abstract representation of obtaining G_p^H from historical solar PV power generation records. Solar irradiance curves are extracted from estimation of the hourly global solar irradiation based on numerical weather predictions.

After assigning all measurements to their corresponding classes, the probabilities for each hour of the day are calculated for the entire dataset based on the classes that share the same hour label. As a result, the matrix G_p^H will have dimensions equal to the number

of hours in a day and the number of classes, classes, which are set to 100 in this case. Then, the final probability matrices are constructed as:

$$\mathcal{G}_p^{\mathcal{F}} + \mathcal{G}_p^{\mathcal{H}} = \mathcal{G}_p^{\mathcal{T}}, \quad \mathcal{G}_p^{\mathcal{T}} \in \mathbb{R}^{\mathcal{R} \times \mathcal{D}}, \quad (4.22)$$

$$D_p^{\mathcal{F}} + D_p^{\mathcal{H}} = D_p^{\mathcal{T}}, \quad D_p^{\mathcal{T}} \in \mathbb{R}^{\mathcal{R} \times \mathcal{D}}, \quad (4.23)$$

where $\mathcal{G}_p^{\mathcal{T}}$ and $D_p^{\mathcal{T}}$, are total probabilities for production and demand ranges, respectively by considering both historical and forecasted values. Whenever probabilities are accumulated, an averaging operation is also performed to ensure that the total probability sum always remains equal to one.

Scenarios are generated based on combinations of paired power generation and consumption values. To achieve this, the Cumulative Distribution Function (CDF) for each hour of the day is derived from the $\mathcal{G}_p^{\mathcal{T}}$ and $D_p^{\mathcal{T}}$ probability matrices. For each CDF curve, the probability range is equally divided into \mathcal{S} sections, where \mathcal{S} represents number of scenarios. Then, one random value is generated for each section, and based on these values, the corresponding points are selected for each time step in the control horizon. Assuming $\mathcal{S} = 100$ for each hour of the day, there are 100 power generation ($G_h^{\mathcal{S}}$) and demand ($D_h^{\mathcal{S}}$) values, where $h \in [0, 23]$ and $\mathcal{S} \in [1, \mathcal{S}]$. Finally, the daily consumption and generation profiles are generated by combining $G_h^{\mathcal{S}}$ and $D_h^{\mathcal{S}}$ as:

$$\mathfrak{S} = \begin{bmatrix} (G_0^1, D_0^1) & \cdots & (G_0^{\mathcal{S}}, D_0^{\mathcal{S}}) \\ \vdots & \ddots & \vdots \\ (G_{h-1}^1, D_{h-1}^1) & \cdots & (G_{h-1}^{\mathcal{S}}, D_{h-1}^{\mathcal{S}}) \end{bmatrix}_{h \times \mathcal{S}}, \quad (4.24)$$

where \mathfrak{S} is the matrix of all generated scenarios. Fig. 4.5 illustrates the described procedure.

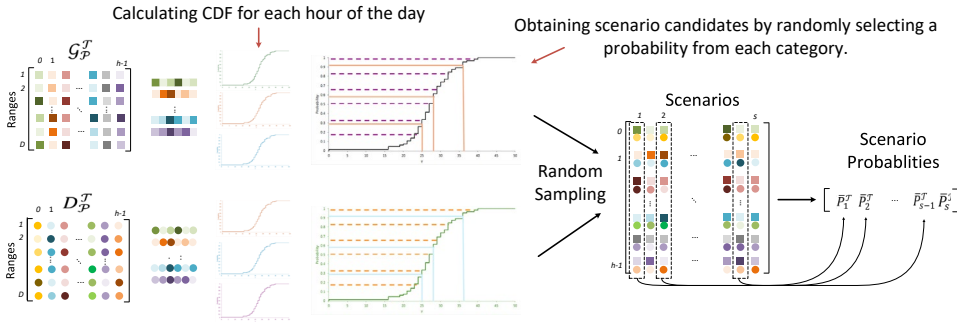


Figure 4.5 Daily scenario generation based on calculated probability matrices and random candidate selection.

Also, since combined generation and consumption values ($G_{h-1}^{\mathcal{S}}, D_{h-1}^{\mathcal{S}}$) have different probabilities, the probability for each pair is considered as multiplication of each individual probability. For calculating the total probability of each scenario ($\bar{P}_s^{\mathcal{T}}$), the probability of each individual hour is accumulated and then averaged. These probabilities are stored in a scenario's probability matrix:

$$P_S^{\mathcal{T}} = [\bar{P}_1^{\mathcal{T}}, \bar{P}_2^{\mathcal{T}}, \bar{P}_3^{\mathcal{T}}, \dots, \bar{P}_S^{\mathcal{T}}], \quad P_S^{\mathcal{T}} \in \mathbb{R}^{1 \times \mathcal{S}}. \quad (4.25)$$

In this method, the main optimization problem consists of sub-optimization problems for each scenario. In other words, the optimal solution is the one that minimizes the defined cost function while considering all scenarios. However, this does not guarantee

that the solution is also optimal for each individual scenario. A general optimization problem is defined as:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^s \mathcal{F}_i^T \cdot X_i \cdot \dot{P}_i^T \\ \text{s.t.} \quad & X_{i,min} \leq X_i \leq X_{i,max} \end{aligned} \quad 4.26)$$

where \mathcal{F}_i is the vector of optimization variables, X_i represents the vector of optimization factors, and \dot{P}_i^T denotes the probability of each scenario's occurring within the corresponding optimization horizon. Additionally, s represents the number of selected candidate scenarios. To manage the complexity of the optimization process, only ten most probable scenarios are selected from the generated set after sorting them by probability. It is also worth noting that the probabilities of the selected scenarios are normalized to reflect their relative differences.

Furthermore, the sub-optimization problem for each scenario is defined as:

$$\min \sum_{h=0}^w E_h^{imp} \times ToU_h^{imp} - E_h^{exp} \times ToU_h^{exp}. \quad 4.27)$$

The objectives and constraints for Eq. 4.27) can be represented as:

$$\forall h, i, \quad 0 \leq P_{A \rightarrow B}^i[h], \quad 4.28)$$

$$\forall h, i, 0 \leq P_{pv \rightarrow gr}^i[h] + P_{pv \rightarrow es}^i[h] + P_{pv \rightarrow ld}^i[h] \leq P_{pv,max}, \quad 4.29)$$

$$\forall h, i, \quad 0 \leq P_{gr \rightarrow ld}^i[h] + P_{gr \rightarrow es}^i[h] \leq P_{gr,max} \quad 4.30)$$

$$\forall h, i, \quad 0 \leq P_{es \rightarrow ld}^i[h] + P_{es \rightarrow gr}^i[h] \leq P_{es,max}, \quad 4.31)$$

$$\forall h, i, P_{pv \rightarrow ld}^i[h] + P_{pv \rightarrow gr}^i[h] + P_{pv \rightarrow es}^i[h] = P_{pv}^i[h], \quad 4.32)$$

$$\forall h, i, P_{es \rightarrow ld}^i[h] + P_{gr \rightarrow ld}^i[h] + P_{pv \rightarrow ld}^i[h] = P_{ld}^i[h], \quad 4.33)$$

$$\forall h, i \quad SoC_{min} \leq SoC_{es}^i[h] \leq SoC_{max} \quad 4.34)$$

$$\forall h, i \quad export_{gr}^i[h] + import_{gr}^i[h] \leq 1, \quad 4.35)$$

$$\forall h, i \quad charge_{es}^i[h] + discharge_{es}^i[h] \leq 1, \quad 4.36)$$

where $E_h^{imp} = (P_{gr \rightarrow ld}[h] + P_{gr \rightarrow es}[h]) \times t$ is the total imported energy for each hour of the day from the grid, and $E_h^{exp} = (P_{pv \rightarrow gr}[h] + P_{es \rightarrow gr}[h]) \times t$ represents the total amount of net energy exported to the grid during hour (h). Also, ToU_h^{imp} , and ToU_h^{exp} denote the time-of-use tariffs for imported and exported energy, respectively. The notation $P_{A \rightarrow B}^i[h]$ represents power flow from point A to point B, and A, B $\in \{pv, gr, es\}$.

Eq. 4.28) ensures that all power flows are non-negative. Furthermore, $P_{pv \rightarrow gr}^i[h]$, $P_{pv \rightarrow es}^i[h]$, and $P_{pv \rightarrow ld}^i[h]$ represent power flow from PV to grid, energy storage, and load, respectively. The notation $P_{pv,max}$ denotes the maximum allowable power output from the PV system, constrained by the PV system size and power electronics limitations. $P_{gr \rightarrow ld}^i[h]$, and $P_{gr \rightarrow es}^i[h]$, represent power flow from grid to load and battery storage,

respectively, and $P_{gr,max}$ is the maximum allowable power exchange with the grid. $P_{es \rightarrow ld}^i[h]$, and $P_{es \rightarrow gr}^i[h]$ represent power flow from energy storage to load and grid, respectively, and $P_{es,max}$ is the maximum charge/discharge power of the battery storage unit.

Eq. 4.32) and 4.33) are equality constraints ensuring that the optimization algorithm satisfies demand and utilizes all available solar PV power under all conditions. Here, $P_{pv}^i[h]$ and $P_{ld}^i[h]$ represent the generated and demanded power, at time h , respectively. In Eq. 4.34), SoC_{min} , SoC_{max} define, respectively, the minimum and maximum allowable battery SoC levels. Finally, Eq. 4.35) and 4.36) prevent the optimization algorithm from generating infeasible solutions. For example, importing and exporting power to the grid simultaneously is physically impossible. Therefore, $export_{gr}^i[h]$, and $import_{gr}^i[h]$ are Boolean values, that enforce this constraint. A similar logic applies to energy storage, where $charge_{es}^i[h]$, and $discharge_{es}^i[h]$ are Boolean variables indicating the charging or discharging state of the battery. If the battery is charging $charge_{es}^i[h] = 1$, and otherwise the $discharge_{es}^i[h] = 1$.

This optimization problem is formulated as a MILP problem. The problem has been implemented using the Pyomo optimization framework and solved using the Gurobi solver.

4.4 Performance Comparison

4.4.1 Key Performance Indicators (KPIs)

Various factors can be considered when evaluating the performance of home EMSs. Depending on optimization goals and problem formulation, these factors may include self-consumption ratio, electricity costs, energy storage utilization or charge/discharge cycles, demand response quality, energy conversion losses, etc. In this thesis, the focus is on two key performance metrics: electricity costs as the primary performance indicator and building self-consumption ratio ($\delta \in [0,100]$ %) as the secondary factor. However, since the optimization problem is formulated solely based on minimizing the energy bill, δ serves only as a performance measurement metric and does not influence the optimization process. Yet, there is a possibility to redefine the optimization problem, and formulate it based on batch factors. In this case, a multi objective optimization function must be solved based on evolutionary algorithm such as genetic algorithm, or PSP methods.

Considering, the energy bill minimization, as the main goal, annual energy bill (\mathcal{A}_{bill}) is defined as:

$$\mathcal{A}_{bill} = \sum_{d=1}^{365} \sum_{h=0}^w E_{h,d}^{imp} \times ToU_{h,d}^{imp} - E_{h,d}^{exp} \times ToU_{h,d}^{exp} . \quad 4.37)$$

where, $E_{h,d}^{imp}$ and $E_{h,d}^{exp}$ represents the amount of total energy the house has been imported and exported from grid or to the grid, respectively. $ToU_{h,d}^{imp}$ and $ToU_{h,d}^{exp}$ represents the import and export electricity tariffs, respectively. Also, the annual self-sufficiency ratio is defined as:

$$\sigma = \frac{\sum_{d=1}^{365} \sum_{h=0}^{23} E_{pv \rightarrow ld}[h, d] + E_{pv \rightarrow es \rightarrow ld}[h, d]}{E_{pv}[h, d]} , \quad 4.38)$$

where $E_{pv \rightarrow ld} = P_{pv \rightarrow ld} \times t$, $E_{pv} = P_{pv} \times t$, and $E_{pv \rightarrow es \rightarrow ld} = P_{pv \rightarrow es \rightarrow ld} \times t$. The notation $E_{pv \rightarrow es \rightarrow ld}$ represents the amount of energy generated by the solar PV system, stored in

the energy storage system, and later delivered to the load. This term is often overlooked in literature, where the self-consumption ratio is typically calculated only by considering the real-time power delivery from the PV system to the load.

The other important factor to consider, is energy autonomy ratio. This factor, quantifies, the undependability of a house to the grid. This means, houses without any on-site renewable energy generation units, are 0 % energy independence, regardless of the size of home battery they might have. In regions far from equator, due to the short days in the wintertime and limited solar energy generation, it is technically infeasible to reach full energy autonomy. So, regardless of the installed on-site solar PV size, this factor has a saturation point for each region. The energy autonomy (\check{e}), is calculated as:

$$\check{e} = \sum_{d=1}^{365} \frac{E_{demand}(d) - E_{import}(d)}{E_{demand}(d)}, \quad 4.39)$$

where E_{import} shows the total amount of imported energy from grid, and E_{demand} is the total amount of daily energy demand. Fig. 4.6 visualize the relationship between these three factors, considering various scenarios.

Annual Energy Generation = Annual Energy Consumption

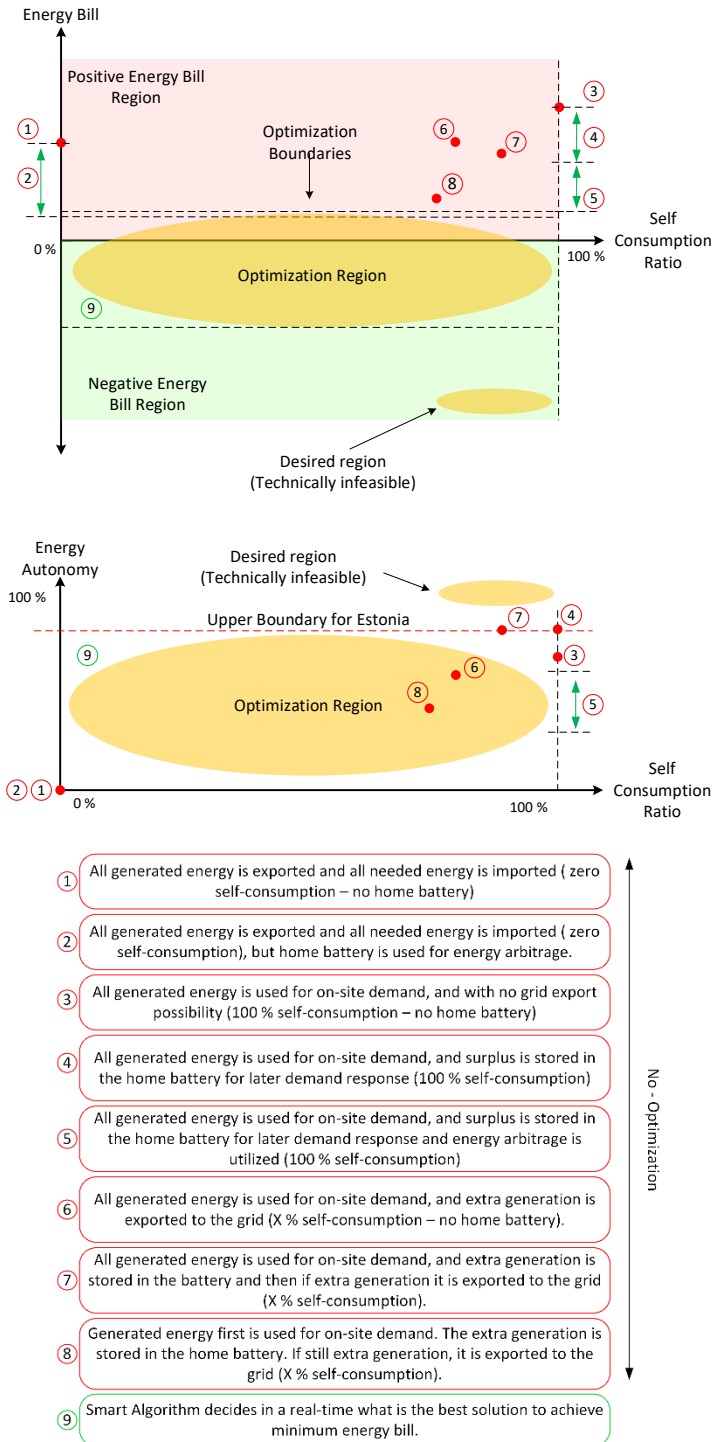


Figure 4.6 Energy bill, Self consumption, and energy autonomy ration correlations, applying different scenarios.

The simplest approach involves exporting all generated energy to the grid while importing all household demand (Scenario 1), resulting in zero self-consumption. This Scenario 1 is not a reasonable strategy and is only considered for comparison purposes. A marginal improvement introduces a home battery used solely for energy arbitrage. This means charging from the grid during low-tariff periods and discharging during high-tariff periods (Scenario 2). This scenario is also not a reasonable strategy to apply but making calculation based by only considering energy arbitrage, gives beneficial insights about home battery applications even when there are no solar PV installations. More restrictive configurations limit the PV system to cover only on-site demand with no provision for grid export or storage, leading to wasted surplus generation (Scenario 3). Scenario 3 is possible to happen in residential buildings without home battery installations, and without permission to export surplus energy to the grid. In these cases, the curtailment strategies must be applied. If home battery is integrated, then this extra energy can be supplied and stored in the home battery for later use. In Scenario 4, excess energy is stored but cannot be exported, and in Scenario 5, energy arbitrage is incorporated into the battery usage strategy, enabling optimized demand shifting under a zero-export constraint.

Grid-interactive strategies without home battery (Scenario 6) prioritize self-consumption and allow the export of surplus energy. This is a common setup in residential systems without batteries. More advanced configurations combine local consumption, battery storage, and conditional grid export. In Scenario 7, energy is first consumed locally, then stored in the battery, and any remaining surplus is exported to the grid. Scenario 8 builds upon this by giving priority to both self-consumption and energy storage, while also enabling export if battery capacity is exceeded. Among all considered configurations, Scenario 9 employs a smart EMS solution capable of real-time decision-making based on current generation, load demand, battery state-of-charge, and electricity tariffs. This dynamic control algorithm continuously evaluates the optimal energy flow to minimize the overall energy cost.

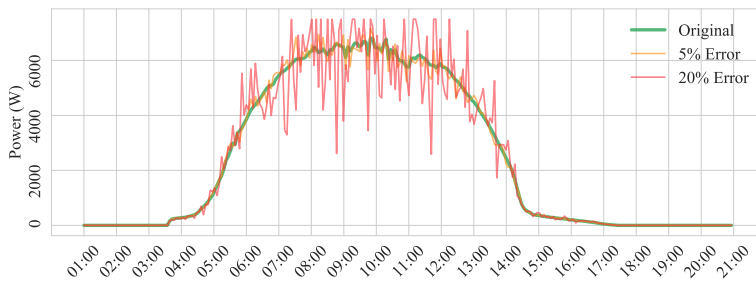
Table 4.1 Performance comparison between various EMS scenarios.

EMS Scenarios	PV*	HB*	OP*	SA*	GE*	EAR*	SFR* (%)	AEB* (€)	TIEG* (kWh)	TEEG* (kWh)
Scenario 1	✓	✗	✗	✗	✓	✗	0	28.03	3388.23	4210.42
Scenario 2	✓	✓	✗	✗	✓	✓	0	-44.05	3388.23	4210.42
Scenario 3	✓	✗	✗	✗	✗	✗	34.74	209.41	2056.47	0
Scenario 4	✓	✓	✗	✗	✗	✗	46.74	148.99	2389.86	0
Scenario 5	✓	✓	✗	✗	✗	✓	34.74	176.47	2389.86	0
Scenario 6	✓	✗	✗	✗	✓	✗	31.74	84.05	2051.97	2874.15
Scenario 7	✓	✓	✗	✗	✓	✗	46.74	71.29	2389.86	2065.21
Scenario 8	✓	✓	Rule-based	✗	✓	✓	51.26	60.91	1651.97	2474.15
Deterministic	✓	✓	Deterministic + Rule-based	✓	✓	✓	51.25	-30.97	2052.95	1297.40
Stochastic	✓	✓	Probabilistic	✓	✓	✓	71.10	-97.48	1721.38	1170.78
Ideal forecasting	✓	✓	Deterministic	✗	✓	✓	85.44	-163.85	1246.82	2224.71
Conventional home	✗	✗	✗	✗	✗	✗	0	331.01	3388.23	0

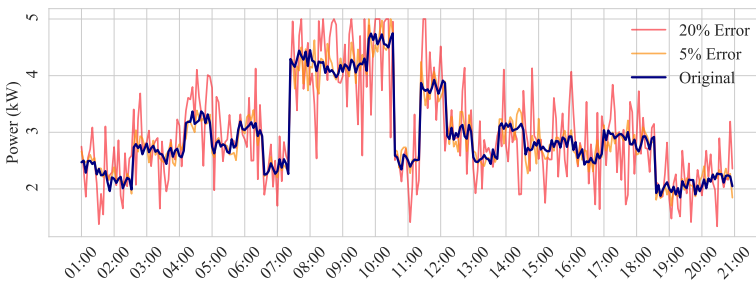
*HB: Home Battery, OP: Optimizations, SA: Smart Algorithm, GE: Grid Export, EAR: Energy Arbitrage, SSR: PV Self-Sufficiency Ratio, AEB: Annual Electricity Bill, TIEG: Total Imported Energy from the Grid, TEEG: Total Exported Energy to the Grid.

From a feasibility and practical deployment perspective, Scenarios 7 through 9 reflect the most realistic and beneficial strategies, particularly in modern smart homes equipped with PV systems, home battery, dynamic tariffs, and advanced EMS capabilities. Table 4.1 compares these scenarios and proposed solutions performances for different parameters. It must be mentioned, the developed solutions in this thesis, can be classified as a scenario 9. So, in the Table 4.1, instead of scenario 9, each specific method is mentioned and considered. Also, an ideal forecasting case is included in the table to give insights about the best possible achievement for a setup.

As one can notice in Table 4.1, the closest performance belongs to the stochastic optimization case. Deterministic scenario fails to compete with stochastic solutions, due to overlooking inherent uncertainty in the system. Rule-based solutions are simple to apply but may not be as effective as complicated methods.



(a)

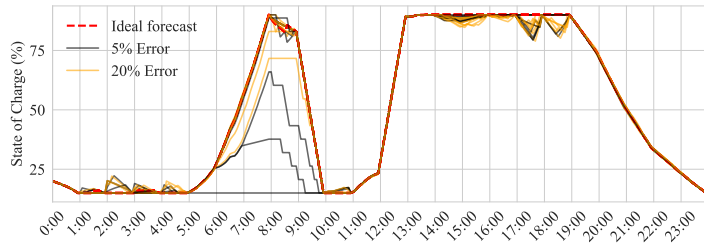


(b)

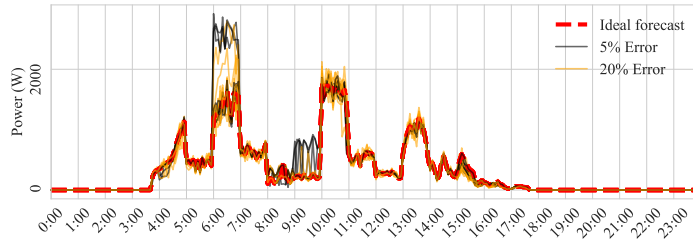
Figure 4.7 Error injections to PV and load profiles. (a) PV profile fluctuations with error injections. (b) Load profile fluctuations with error injections.

4.4.2 Sensitivity Analysis

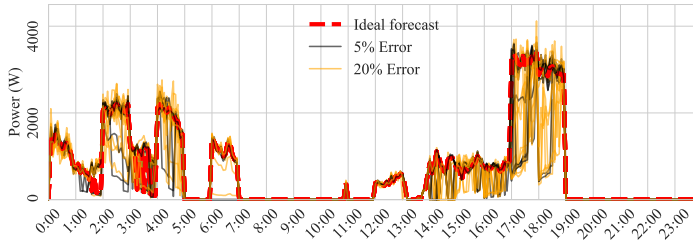
Due to the intermittent nature of renewable resources and uncertainties related to energy consumption profiles within dwellings, optimizing energy flow without accounting for these natural uncertainties will lead to deviations from ideal operational assumptions. To analyze the effects of uncertainties on the system's operational logic, Monte Carlo simulation has been performed. To this end, random errors have been introduced into the PV and load profiles. These errors are assumed to follow a Gaussian distribution with a mean value of zero and standard deviations of 0.05 and 0.2, corresponding to error levels of 5% and 20%, respectively. Fig. 4.7(a) and Fig. 4.7(b) display the error injections to the daily PV and load profiles.



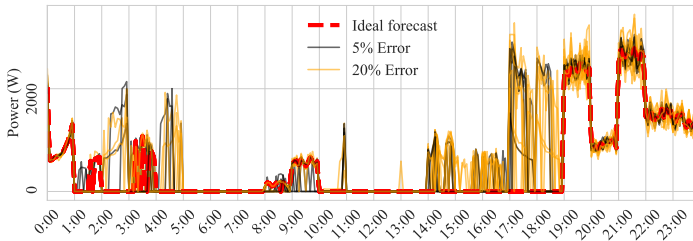
(a)



(b)



(c)



(d)

Figure 4.8 System variables deviations from ideal condition in presence of error injection to PV and Load profiles. (a) ESS SoC level deviations from ideal condition. (b) PV to load delivered power deviations from ideal condition. (c) Grid to load delivered power deviations from ideal condition. (d) ESS to load delivered power deviations from ideal condition.

The outcomes of these random error injections have been analyzed after 100 iterations. Deviations in the system's variables are depicted in Fig. 4.8(a)–(d). It is observed that the parameter most affected among the demand response elements is the power flow from home battery to the load. Conversely, the home battery's SoC profile shows the least deviation from ideal performance, despite these variations. This phenomenon can be attributed to the fact that deviations do not persist for long periods, and the impact of errors on energy flow optimization has a short-lived effect. Yet, at certain times, the EMS's decisions regarding energy flow supervision and demand response follow

entirely different scenarios, especially in cases of 20% error injection. This leads the algorithm to converge to different minimum, indicating that the deterministic EMS solution's robustness should be examined more thoroughly.

4.5 Summary

In this chapter, the building EMS has been introduced and formulated, and multiple EMS strategies have been developed, including heuristic, deterministic, and probabilistic methods. The performance comparison between these approaches reveals that, from an implementation perspective, heuristic or rule-based methods are the most straightforward to apply; however, they typically fall short in delivering high-performance outcomes.

On the other hand, probabilistic methods, while the most challenging to implement due to their computational complexity and requirements for HPPUs, offer the most robust and reliable performance. This is largely due to their ability to incorporate uncertainty into the optimization process, resulting in more accurate and resilient decision-making.

Deterministic methods provide a middle ground, balancing computational complexity and performance between heuristic and probabilistic approaches. These methods can include forecasting techniques to estimate near-future system states and make decisions with increased confidence. However, they do not account for the inherent uncertainties within the system. In probabilistic and deterministic methods, to ensure uninterrupted power delivery, it is necessary to implement an additional layer of rule-based or high-resolution control mechanisms. These layers are responsible for overriding or adjusting the control signals generated by the optimization algorithm in scenarios where unforeseen changes in power generation or demand could lead to failure in power delivery from the dedicated source.

This chapter is based on paper III, paper V, paper VI, and [48] and addresses research task number 3 and hypothesis number 3. The main contribution is development of a two home EMS solution for deterministic and stochastic resource optimization and allocation in the residential nZEBs. Also, sensitivity analysis was conducted to investigate the sensitivity of EMS performances related to forecasting errors.

5 Size Optimization and Cost Justification

Infrastructure size optimization is a critical step toward developing sustainable and cost-effective solutions, particularly in residential applications. Poorly optimized systems often result in high upfront costs and extended payback periods, deterring potential adopters. The absence of region-based and case-based techno-economic models has left end-users uncertain about the profitability and long-term benefits of implementing home battery and EMSs. Thus, a comprehensive, user-specific model that integrates technical, economic, and contextual factors can provide clearer insights for both customers, domain experts, and policymakers. To address these requirements, a comprehensive techno-economic model has been developed in this thesis to evaluate the influence of various parameters on system profitability and to estimate the payback period based on user-specific input factors.

5.1 Techno-Economic Model Development

The developed model incorporates a physics-based battery SoH estimator, a linear PV degradation model, and an online tariff extraction module. In addition, it accounts for key economic variables such as energy market trends, inflation rates, and long-term economic projections.

5.1.1 Battery Degradation Model

Many previous studies have largely overlooked the impact of battery degradation on model performance. While this simplification may be justifiable for systems that are inherently unprofitable, it becomes a critical limitation in models with the potential for cost-effectiveness. In such cases, accounting for the aging effects of lithium-ion batteries is presumed essential. The rate of battery degradation is highly dependent on the usage strategy, meaning that different operational approaches can significantly influence battery longevity. Consequently, incorporating degradation-aware optimization can lead to strategies that not only extend battery life but also enhance the overall economic performance of the system.

Capacity fading is the most significant drawback of Li-ion battery cell technology. This parameter refers to remained cell capacity in comparison with nominal cell capacity. This phenomenon is specified with two major factors: degradation due to time pass and fading due to the amount of processed charge [49]. As a result, total capacity degradation $C_{Fd} \in [0,100]$ can be obtained as:

$$C_{Fd} = C_{cal} + \frac{C_{cycle}}{C_{nm}} \times 100, \quad 5.1)$$

where $C_{cal} \in [0,100]$ represents the calendar capacity loss due to battery stock or long-term relaxation and $C_{cycle} \in [0, C_{nm}]$ is the capacity degradation resulting from the amount of the processed charge during cells recharge/discharge cycles and $C_{nm} \in (0, 15]$ kWh is the nominal capacity of the fresh battery cell. Finally, by considering both degradation factors, the SoH $\in [0,100]\%$ level of the energy storage can be formulated as:

$$SoH = 100 - C_{Fd}. \quad 5.2)$$

For fresh batteries C_{Fd} is zero and SoH is equal to 100%. Thus, SoH equals to zero indicates that all capacity of the energy storage is decayed, and the battery is no longer capable of storing and delivering energy.

In the developed energy storage fading model, calendar capacity loss is applied using the superposition rule and is calculated as [50]:

$$C_{cal,fade} = (0.019 \times SoC^{0.823} + 0.5195) \times (3.258 \times 10^{-9} \times T^{5.087} + 0.295) \times m^{0.8}, \quad 5.3)$$

where $C_{cal,fade} \in [0,100]$ represents the percentage of total calendric capacity loss over the course of m months, and T denotes the ambient temperature in degrees centigrade, and SoC refers to the cells' available charge level, expressed in percentage, during batteries' stock period.

To calculate the calendar loss during energy storage operation, Eq. 5.3) should be reformed. For simplification, the SoC value is approximated to be constant at 1-hour intervals. As a result, the SoC value is averaged for the corresponding time slot, represented by $SoC_{1h,avg} \in [0,100]$. Given that the cell temperature is maintained at 25 °C, the capacity fading ratio for 1 hour period represented by $C_{cal,1h} \in [0,100]$ can be computed as follows:

$$C_{cal,1h} = (6.6148 \times SoC_{1h,avg} + 4.6404) \times 10^{-6}. \quad 5.4)$$

In Eq. (5.4), $C_{cal,1h}$ represents the percentage of the capacity loss during each hour of storing battery cells in the mentioned condition and $SoC_{1h,avg}$ is the average of SoC levels during each hour. So, it can be concluded that a higher SoC ratio accelerates calendar ageing and for a long-term stock, cells should not be fully charged. Finally, the total calendric capacity fade represented by C_{cal} can be linearly expressed as:

$$C_{cal} = \sum_{\tau=1}^h C_{cal,1h}[\tau], \quad 5.5)$$

where h is the total hours of energy storage operation and relaxing times.

To calculate the capacity fading of cells due to charge circulations, the model proposed in [51] has been applied. In this empirical model, energy storage degradation is calculated based on average SoC level and its deviation from the average value during each charge/discharge event. The total cyclic degradation is compiled as:

$$C_{cycle} = \sum_{i=1}^e ((k_{s1} SoC_{dev,i} e^{(k_{s2} * SoC_{avg,i})} + k_{s3} e^{k_{s4} * SoC_{dev,i}}) e^{\left(\frac{-E_a}{R} \left(\frac{1}{T_i} - \frac{1}{T_{ref}}\right)\right)}) Q_i, \quad 5.6)$$

where C_{cycle} is calculated by accumulating all recharge and discharge cycles i . Q_i is the amount of processed charge during each recharge/discharge cycle which is function of cell's current and time duration of passed current. Here, $k_{s1}, k_{s2}, k_{s3}, k_{s4}$ are the parameters of the empirical capacity fading model. E_a is the activation energy, R is the gas constant, T_i is the cell temperature at the i th charge or discharge moment, T_{ref} is a reference temperature, both in Kelvin, and e is the total number of recharge and discharge cycles. The numerical values for the mentioned parameters are collated in Table IV. In Eq. 5.6), SoC_{avg} is average amount of SoCs and SoC_{dev} is normalized standard deviation of SoC from SoC_{avg} during each charge/discharge operation. For instance, if a full-charged cell is completely discharged in one cycle, then both SoC_{avg} and SoC_{dev} will be obtained as 50%. SoC_{avg} and SoC_{dev} have been formulated as:

$$SoC_{avg} = \frac{1}{\Delta Q_m} \int_{Q_{m-1}}^{Q_m} SoC(Q) dQ, \quad 5.7)$$

$$SoC_{dev} = \sqrt{\left| \frac{3}{\Delta Q_m} \int_{Q_{m-1}}^{Q_m} (SoC(Q) - SoC_{avg})^2 dQ \right|}, \quad 5.8)$$

where Q_m is the final amount of processed charge, Q_{m-1} is the initial amount of processed charge before starting of energy circulation through battery cells, and ΔQ_m is the absolute amount of charge processed in each operation cycle. It should be mentioned that this model holds accuracy only under conditions where the cells are persistently operating at temperatures above 25 °C.

Table 5.1 Circular capacity fading model's constant parameters.

Parameter	Value	Unit
k_{s1}	-4.092e-4	-
k_{s2}	-2.167	-
k_{s3}	1.408e-5	-
k_{s4}	6.13	-
E_a	78.06	k.mol/J
R	8.314	J/k.mol
T_{ref}	298.15	K
T_i	303.15	K

5.1.2 Techno-Economic Model

Table 5.2 lists the considered financial and certain technical parameters that impact the economic performance of the model.

Table 5.2 Financial and technical parameters.

Parameter	Value	Unit	Symbol
Li-ion BESS cost	750	€/kWh	U_{BESS}
BESS annual price reduction [52]	12%	-	Ω_{BESS}
BESS value/size reduction	3%	Per kWh	$\Omega_{BESS/size}$
PV cost [53]	1300	€/kW	U_{PV}
PV value/size reduction	4%	Per kW	$\Omega_{PV/size}$
PV lifetime [54]	30	Year	Ψ_{PV}
ER cost [55]	250	€/kW	U_{ER}
ER lifetime [54]	15	Year	Ψ_{ER}
ER efficiency [55]	98	%	Ω_{INF}
ER capacity	2.5-20	kW	P_{max}
EMS cost	200	€/unit	U_{EMS}
Load annual increase rate	3%	Per year	Ω_L
Utility tariffs inflation rate	5%	Per year	Ω_{UT}
Annual inflation rate	4%	Per year	Ω_{AN}
Interest rate	3.5%	Per year	Ω_{IN}
Feed-in tariff coefficient	0.8	-	χ
Maintenance cost [56]	1% of setup cost	€/year	U_M

Payback time is the duration required to recover an investment. The primary factors in calculating the payback time are the initial installation Capital Expenditure (CapEx), inflation, and the economic performance of the system during operation.

There are two methods to calculate this parameter: simple and discounted. The annual inflation rate is a crucial factor in computing returns, rendering the simple payback calculation method imprecise due to its exclusion of this rate. Consequently, the discounted payback time ϑ , which incorporates inflation rates in the return calculation, is more accurate and is computed as:

$$\vartheta = \ln\left(\frac{1}{1 - \frac{\xi_c \Omega_{AN}}{\pi_Y}}\right) \div \ln(1 + \Omega_{AN}), \quad 5.9)$$

where π_Y represents the total annual cash flow in the system, ξ_c denotes the initial investment, and Ω_{AN} signifies the annual inflation. However, due to the impact of system fading on performance over time, causing π_Y to vary each operational year, modifications to the conventional discounted payback formula are necessary to yield accurate results. Consequently, Eq. 5.9) is modified as:

$$\vartheta_u = -\xi_c - \xi_A + \sum_{k=1}^m \frac{\varphi_k - \Phi_k}{(1 + \frac{\Omega_{AN}}{12})^k}, \quad 5.10)$$

where ϑ_u is the remaining investment amount for compensation, ξ_A is the annual operating cost of the system, and φ_k is the utility bill in the month number k if the user has not installed any solar or energy storage. For instance, considering a corresponding k_{th} month to be 30 days, $\varphi_k = \sum_{j=1}^{24*30} E_{Load,j} \times U_{ToU,j} \cdot \Phi_k$ represents monthly net utility bill after deploying the PV and home battery setup and calculated as follows:

$$\Phi_k = \sum_{j=1}^{24*30} [(E_{g,j} - \gamma E_{f,j})U_{ToU,j} + E_{Ex,j}C_{Ex}] + C_{fx}. \quad 5.11)$$

In Eq. 5.10), the term $\sum_{k=1}^m \frac{\varphi_k - \Phi_k}{(1 + \frac{\Omega_{AN}}{12})^k} - \xi_A$, represents the net present value of the profit the user accrues after equipping the house with the PV and home battery setup. ξ_c is compiled as:

$$\xi_c = U_{BESS}C_{nm}(1 - \Omega_{BESS/size}C_{nm}) + U_{PV}\theta_{PV}(1 - \Omega_{PV/size}\theta_{PV}) + U_{ER}P_{max} + U_{EMU}, \quad 5.12)$$

where θ_{PV} is the solar system size (kWp), U_* is the unit price of installing each of the PV, home battery, ER, and EMU. P_{max} is the selected EnergyR size in kW, and $\Omega_{*/size}$ represents the cost/size reduction rate for both energy storage and PV panels. Also, ξ_A is compiled as:

$$\xi_A = \xi_c U_M (1 + \Omega_{An})^y + \kappa_{BESS} U_{BESS} (1 + \Omega_{An} - \Omega_{BESS})^{\Psi_{BESS}} + \kappa_{ER} U_{ER} (1 + \Omega_{An})^{\Psi_{ER}}, \quad 5.13)$$

where y represents the running age of the setup in years, and $U_M \in [0,1]$ represents the annual maintenance cost ration, and here is assumed to be 1% of ξ_c . κ_{ER} and κ_{BESS} denote the number of times the EnergyR and energy storage have been replaced respectively during setup run, Ψ_* indicates the year in which the replacement occurred, and Ω_{BESS} represents the annual price reduction in the market for the energy storage technology. Based on Eq. 5.10), the payback time is the first month in which the ϑ_u turns to a positive value. Algorithm 5.1 summarizes the calculation procedure in the highlighted section.

Algorithm 5.1 *Economical parameters calculation*

```
1:   Input:  $\vec{E}_g, \vec{E}_f, \vec{U}_{ToU}, \vec{PV}_{size}, \vec{BESS}_{size}, \vec{ER}_{size}$ , Table VI parm.

2:   for  $v, s, r$  in  $\vec{PV}_{size}, \vec{BESS}_{size}, \vec{ER}_{size}$  do
3:     Calculate  $\xi_c[v, s, r]$  based on Eq. (35).
4:     Calculate  $\xi_A[v, s, r]$  based on Eq. (36).
5:     for  $k$  in  $m$  do
6:       Calculate  $\Phi_k[v, s, r]$  based on Eq. (34):
7:         
$$\Phi_k \leftarrow \sum_{j=1}^{24 \times 30} [\vec{E}_{g,v,s,r,j} - \forall E_{f,v,s,r,j}] U_{ToU,j} + E_{Ex,v,s,r,j} C_{Ex} + C_{fx}$$

8:       Calculate  $\vartheta_{u,k}[v, s, r]$  based on Eq.(33):
9:         
$$\vartheta_{u,k} \leftarrow -\xi_c[v, s, r] - \xi_A[v, s, r] + \sum_k \Phi_k$$

10:      if  $\vartheta_{u,k} \geq 0$  then
11:        payback[v, s, r] = k
12:      end if
13:    end for
14:    Calculate  $\delta[v, s, r]$  based on Eq. (30)
15:    Calculate  $\varepsilon[v, s, r]$  based on Eq. (31)
16:  end for
17:  for  $v$  in  $\vec{PV}_{size}$  do
18:    Select  $s' \in s, r' \in r$  which satisfies:
19:      
$$\forall (s, r), \vartheta_{u,k}[v, s', r'] \leq \vartheta_{u,k}[v, s, r]$$

20:  end for
21:  Return: optimal setup values,  $\vec{\vartheta}, \vec{\delta}, \vec{\varepsilon}$ 
```

5.1.3 Results

Results show that, energy storage capacity degradation significantly affects the system's performance in terms of both revenue acquisition and the quality of demand response. In all analyses. The pivotal role of battery nominal size in influencing the rate of cells aging is evident from the simulation results. Fig. 5.1 shows the SoH level of home battery for all simulated configurations, with Fig. 5.1(a) illustrating a more gradual degradation in larger battery packs compared to smaller ones. This can be attributed to the slower increase in the number of charge and discharge cycles experienced by larger battery packs during operation. Nonetheless, the aging rate of the batteries are multifactorial, being affected not only by the cycle number but also by other elements such as cell temperature, DoD level, and cell current during charges and discharges. In practical experiments, it has been reported that capacity fading exhibits an exponential trend for Li-ion battery cells [57]. Generally, assuming normal operating conditions—encompassing one charge/discharge cycle per day, standard operating temperature, and DoD—the linear annual degradation rate for Li-ion batteries can be estimated to range between 2–4%.

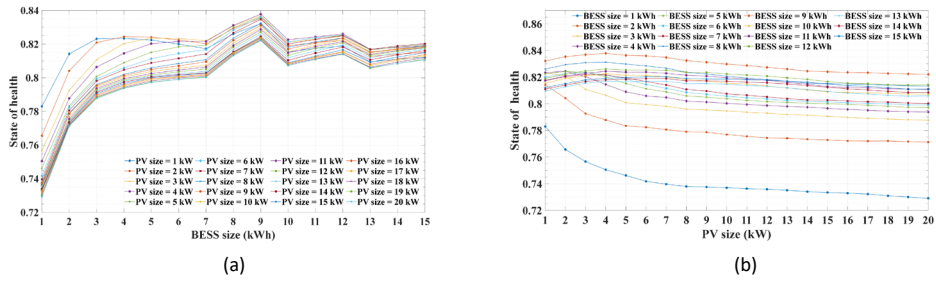


Figure 5.1 Battery degradation ratio for all possible combinations of PV and home battery. (a) Degradation level dependency on BESS size. (b) Degradation level dependency on solar generation setup size.

Moreover, the PV size has a bearing on the battery packs' degradation rate; larger solar power generation can lead to increased battery usage and, consequently, additional degradation. This correlation is demonstrated in Fig. 5.1(b). Comparing Fig. 5.1(a) and Fig. 5.1(b) reveals that the SoH level is predominantly influenced by battery size over PV power capacity. This observation is attributed to the fact that setups with larger solar power generation units tend to export excess energy to the main grid more, thus earning revenue, reducing payback time.

Fig. 5.2 shows investment returns (dotted red line) and other pertinent economic factors for the studied house, illustrating the annual cash flow, capital, maintenance, and replacement costs for a case study house in Estonia. In the initial decade, the owner is primarily responsible for annual maintenance and capital costs. However, in the second decade, replacement costs for BESS and power electronic devices are incurred, typically arising at the 10th and 15th years of operation, respectively. This fee is represented as a step down in the cash revenue graphs, aligning with the replacement costs. In all configurations, if any BESS setup reaches its EoL point before the expiration of its warranted lifetime, it is promptly substituted with a new battery pack.

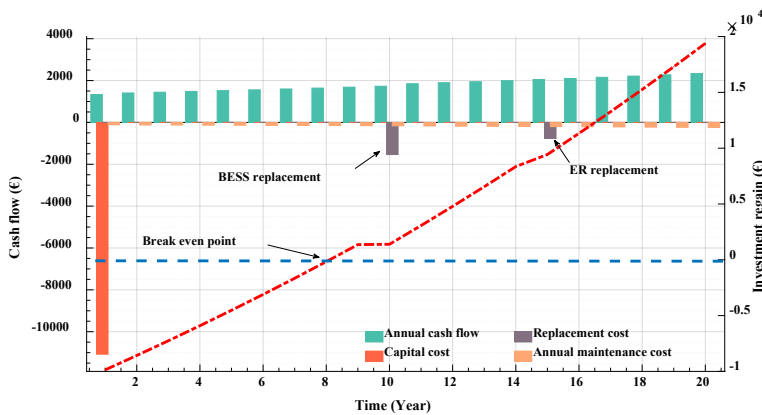
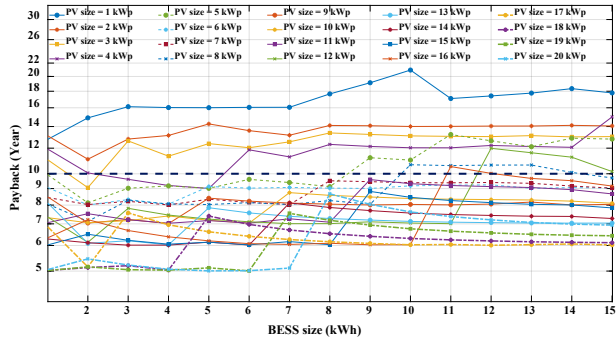


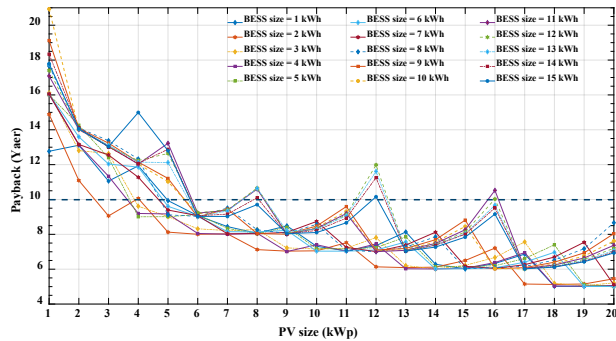
Figure 5.2 Economic performance analysis of the case study house in Estonia.

Fig. 5.3(a) presents a discounted payback time analysis for all PV-home battery combinations. Given the four-dimensional nature of the generated results, the EnergyR dependency is omitted to facilitate clearer visualization, and only the most suitable

EnergyR is selected for each PV-home battery size. From the outcomes, the minimum achievable payback time is observed to be around 5 years. The blue horizontal dashed line in Fig. 5.3(a) and Fig.5.3(b) illustrates the profitability frontier. Here, 10 years is specified in accordance with the lifespan of battery packs. Thus, if the system cannot reach the break-even point within the aforementioned time, it is deemed economically impractical. The results reveal that considering the energy market condition in 2022, choosing the size of PV systems larger than 9 kWp will mostly yield income for the owner.



(a)



(b)

Figure 5.3 Discounted payback calculation of PV-home battery combinations. (a) For each BESS size when the PV size is swept in its range. (b) For each PV size when the BESS size is swept in its range. The Y axis has a logarithmic scale in (b).

Fig. 5.3(b) indicates the dependencies of payback time on BESS size for each specific PV system. This observation demonstrates that, in Estonia, for residential applications, PV-home battery systems with a PV size less than 4.5 kWp are not profitable, even with very small-sized home batteries. Moreover, the impact of PV size on payback time is substantially more pronounced than the impact of home battery size. This is predominantly because, with partial green energy generation, the majority of BESS capacity is allocated to energy arbitrage. Nevertheless, due to battery protection policies and constraints in power electronics, energy arbitrage exhibits lesser dependence on home battery size variations when the PV size is held constant. Consequently, in the present circumstances in Estonia, solely installing home battery in houses without PV systems does not emerge as a feasible solution for long-term use.

Table 5.3 displays the most profitable home battery—EnergyR sizes for each PV arrangement, along with the associated repayment periods and expected income after 20 years of operation. These outcomes are derived from evaluating two distinct scenarios. The first scenario considers the current incentive under the feed-in policy, while the second scenario assumes the termination of the feed-in policy for residential applications. In the latter scenario, surplus solar energy is channelled to the home battery, optimizing system benefits to meet load demands; thus, no electricity is fed back into the utility grid.

Table 5.3 Most profitable combinations' economical results.

PV size (kWp)	FiT	Battery size (kWh)	ER size (kW)	Payback (year)	Revenue (€)
1	0.8 ToU	1	2.5	12.25	2484
	-	1	2.5	17.1	507
2	0.8 ToU	2	2.5	11.09	5603
	-	2	2.5	18.62	841
3	0.8 ToU	2	2.5	9.05	8524
	-	2	2.5	19.15	349
4	0.8 ToU	3	5	9.6	11463
	-	2	5	19.63	106
5	0.8 ToU	4	5	9.1	14528
	-	2	5	> 20	-1097
6	0.8 ToU	4	5	8.06	17779
	-	2	5	> 20	-1822
7	0.8 ToU	5	7.5	8.4	21117
	-	3	7.5	> 20	-2480
8	0.8 ToU	5	7.5	8.03	24556
	-	3	7.5	> 20	-3095
9	0.8 ToU	6	10	8.22	27983
	-	3	10	> 20	-3653
10	0.8 ToU	6	10	7.65	31516
	-	3	10	> 20	-4159
11	0.8 ToU	7	10	7.25	35117
	-	3	10	> 20	-4614
12	0.8 ToU	8	12.5	7.01	38574
	-	4	12.5	> 20	-5014
13	0.8 ToU	9	12.5	7.18	42231
	-	4	12.5	> 20	-5368
14	0.8 ToU	9	12.5	7.69	45956
	-	4	12.5	> 20	-5672
15	0.8 ToU	10	15	8.45	49566
	-	4	15	> 20	-5931
16	0.8 ToU	10	15	6.03	53429
	-	4	15	> 20	-6146
17	0.8 ToU	12	15	6.02	57269
	-	4	15	> 20	-6321
18	0.8 ToU	14	17.5	6.11	61139
	-	5	17.5	> 20	-6456
19	0.8 ToU	15	17.5	6.42	65154
	-	5	17.5	> 20	-6559
20	0.8 ToU	15	20	6.93	69126
	-	5	20	> 20	-6631

5.2 Single-Cell Three-Phase Solution

The payback time results presented in the previous section underscore the importance of delivering cost-effective solutions to enhance the economic viability of EMS integration. In this context, hardware improvements are just as vital as software advancements within such complex systems. The energy router, serving as the hardware core of the proposed EMS platform, plays a pivotal role in enabling diverse operational possibilities and system flexibility.

Simultaneously, the rapid integration of renewable energy sources, has significantly transformed the topology of modern power distribution networks. This transition demands that power systems accommodate a growing proportion of distributed behind-the-meter renewable energy source, introducing new challenges in managing decentralized and localized energy flows. Also, grid congestion and phase imbalances have become more prominent, especially in residential areas, due to the increasing prevalence of home EV charging facilities.

Considering both economic and technical challenges, this thesis introduces a single-cell three-phase EnergyR topology. In this configuration, the EnergyR is designed to interface with all three phases of a residential three-phase grid connection. This capability enables the EnergyR to provide phase balancing services not only within the household electricity network but also to the external grid. Such functionality addresses growing concerns related to phase imbalance, enhancing the stability and efficiency of modern residential power systems.

5.2.1 Single-Cell Three-Phase EnergyR Topology

Fig. 5.4 shows the proposed topology for a Single-Cell Three-Phase (SC-TP) EnergyR and its experimental prototype. In this configuration, a dc bus can interact with all ac phases through an energy router, but not simultaneously. Phase balance can be enhanced by detecting and reducing the power consumption of the phase with the highest demand. Since the power drawn from the three phases in a three-phase connected buildings often varies significantly, reducing the phase imbalance ratio with the proposed solution could lead to substantial economic benefits by eliminating the need for two additional converting cells. Table 5.4 shows the technical specifications for SC-TP energy router.

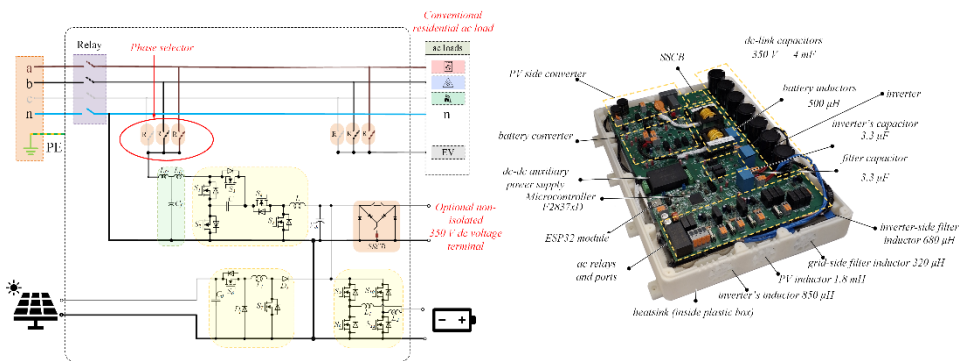


Figure 5.4 Abstract view of the proposed single cell topology connected to three-phase terminal with EV charger, PV, and ES integration and its experimental realization [58].

Table 5.4 EnergyR technical parameters.

Parameters	Value
Rated power	15 kW
Grid and load side ac voltage (RMS)	230 V– 50 HZ
dc-link voltage	350 V
Nominal current of each phase	25 A
Switching frequency	65 kHz
Solar voltage input range	150 - 600 V
ES voltage input range	150 – 330 V
dc-link capacitor	3 mF

5.2.2 Analysed Scenarios and Problem Formulation

To validate the hypothesis and assess the effectiveness of the proposed topology, data on annual load consumption from a residential house located in Tallinn, Estonia, were collected and analyzed. The house is connected to the electricity grid via a three-phase terminal, and the owner has an EV. The data were collected at a 3-second resolution; however, for simplicity and improved visualization, it was averaged and down sampled to a 1-hour resolution. Fig. 5.5 illustrates and compares a snapshot of initial phase imbalances and phase statuses after ideal phase balancing actions. In initial mode, it is evident that phase 3, labeled “L3”, delivers less power compared to the other two phases, indicating that the EV charger should be connected to this phase. However, as observed, severe phase imbalance occurs during EV charging, regardless of which phase the EV is linked.

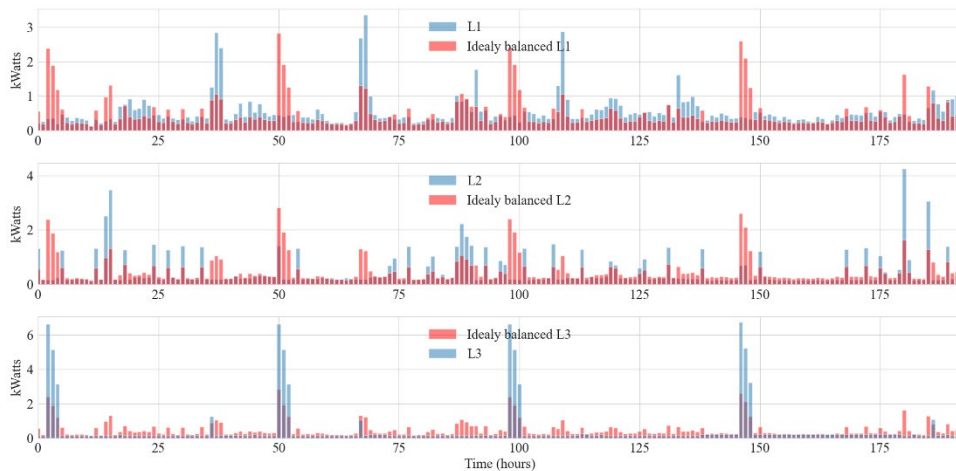


Figure 5.5 Imbalanced and balanced phase power range comparison assuming ideal phase balancing possibilities.

In this thesis, three EnergyR topologies include: Single-Cell Single-Phase (SC-SP), Three-Cell Three-Phase (TC-TP), and SC-TP, are considered and their capability for phase balancing have been compared with each other. The SC-TP EnergyR is the upgraded version of the SC-SP energy router. The 10 kWh Li-Ion battery pack is used as an ES.

Here, the phase imbalance ratio is defined as:

$$UB \% = \sum_{t=0}^T \frac{L_{unbl,t}}{L_{avg,t}} \times \frac{100}{N} \quad 5.14$$

$$L_{unbl,t} = \text{Max} (|L_{avg,t} - L_{1,t}| + |L_{avg,t} - L_{2,t}| + |L_{avg,t} - L_{3,t}|) \quad 5.15$$

$$L_{avg,t} = (L_{1,t} + L_{2,t} + L_{3,t})/3 \quad 5.16$$

where $L_{1,t}, L_{2,t}, L_{3,t}$ is a power consumption for phase 1, 2, 3 at time t, respectively. $L_{avg,t}$ is the average power consumption from all phases in time t and $L_{unbl,t}$ is a maximum power deviation from $L_{avg,t}$ at time t. N is the total number of time steps in which the phase unbalance is calculated. Finally, $UB \%$ is the average phase imbalance. To minimize the $UB\%$ ratio an optimization problem is formulated for the TC-TP mode as:

$$\text{Min } f(x) = \text{Min} (|L''_{avg} - L''_1| + |L''_{avg} - L''_2| + |L''_{avg} - L''_3|) \quad 5.17$$

subject to:

$$L''_1, L''_2, L''_3 \geq 0 \quad 5.18$$

$$PV_{l_1} + PV_{l_2} + PV_{l_3} \leq PV \quad 5.19$$

$$ES_{l_1} + ES_{l_2} + ES_{l_3} \leq ES \quad 5.20$$

$$L_i - PV_{l_i} - ES_{l_i} = L''_i, \quad i = 1, 2, 3 \quad 5.21$$

where $L''_{avg} = (L''_1 + L''_2 + L''_3)/3$ represents the average demand and L''_y is each phase's demand from the electricity grid after allocating available renewable energy resources. Here, y denotes the phase number. PV_{l_y} and ES_{l_y} represents the allocated energy from solar energy production and home battery, respectively. PV and ES denote the total accessible energy from renewable setups and batteries, respectively. In Eq. (5.17), x represents optimization factors, $x = [PV_{l_1}, PV_{l_2}, PV_{l_3}, ES_{l_1}, ES_{l_2}, ES_{l_3}]$. Finally, the home battery SoC level should be updated as:

$$ES_{t+1} = \text{min}(ES_{max}, ES - ES_{l_1} - ES_{l_2} - ES_{l_3} + PV') \quad 5.22$$

where, ES_{t+1} is the home battery, SoC (%) for the next time step, and ES_{max} denotes the maximum energy capacity of the home battery, and PV'_t denotes the remaining generated solar energy after demand responding. It should be mentioned that, in all equations, energy is considered instead of power, since this assumption simplifies the equations.

In SC-TP mode, the optimization problem must be reformulated. Algorithm I outlines the optimization process for SC-TP mode, where the phase with the maximum load demand is identified at each time step. The EnergyR then links this phase to the DC link and injects renewable or stored energy to meet the demand, in the selected phase. In SC-SP mode, since phase exchange is not possible and the EnergyR remains permanently connected to phase "L1," the optimization problem simplifies into a reduced version of Algorithm I, with the index consistently set to 1.

Algorithm 1 SC-TP Energy Optimization Algorithm

```

1: Input: Annual load profiles ( $\vec{l}_1, \vec{l}_2, \vec{l}_3$ ), Annual PV profile (PV).
2: for  $i = 1$  to  $n$  do:
3:    $l_{index, index} = \max(l_{1,i}, l_{2,i}, l_{3,i})$ 
4:    $index = [1, 2, 3]$ 
5:    $x, y = index.drop(index)$ 
6:    $Min f(x) = \min(|l_{avg,i}'' - l_{index,i}''| + |l_{avg,i}'' - l_{x,i}''| + |l_{avg,i}'' - l_{y,i}''|)$ 
7:   Subject to:
8:      $l_{avg,i}'' = avg(l_{index,i}'', l_{x,i}'', l_{y,i}'')$ 
9:      $l_{index,i}'' \geq 0$ 
10:     $l_{index,i}'' = l_{index,i} - PV''_i - ES''_i$ 
11:     $PV''_i \leq PV_i$ 
12:     $ES''_i \leq ES_i$ 
13:     $ES_{i+1} = \min(ES_{max}, ES_i - ES''_i + PV_i - PV''_i)$ 
14:
22: end for
23: Return: optimal values  $PV''_i, ES''_i$ 

```

Fig. 5.6 presents a normalized radar chart comparison across all topologies, measuring variables such as UB%, PV self-consumption ratio, average SoC (%) of home battery, capital costs and average load per phase during an experimental test period. Each category is normalized to its maximum observed value. For example, the average load per phase is highest when the house does not utilize any local renewable energy sources.

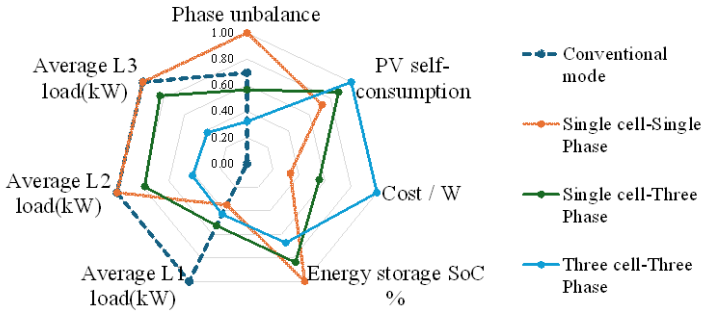


Figure 5.6 Normalized radar chart comparison for all topologies.

Calculations show, in conventional houses, the phase balance ratio is better than in scenarios where the PV setup is connected to only one phase. The greatest phase imbalance is observed in the SC-SP topology, indicating that the presence of PV in this configuration reduces the phase balance ratio. This is because, in such operational modes, the generated renewable energy is only injected into one phase and significantly lowers the demand on that phase alone. Meanwhile, the other phases continue to draw the same demand from the grid and are unable to benefit from PV production or stored energy. Considering that the SC-SP topology is a commonly used option in residential PV setups, this underscores the necessity of proposing novel solutions for improving phase balance ratio.

The TC-TP topology outperforms other configurations in all categories, including PV self-consumption ratio, phase balance ratios, and average home battery's SoC (%) level. However, its capital cost ratio is 1.35 times higher than that of the SC-TP topology, posing a further barrier to adoption in residential buildings, where economic feasibility is a

critical factor. Furthermore, the TC-TP topology can reduce grid interaction across all phases at a similar rate, whereas the SC-TP topology does not achieve the same performance, with the management algorithm primarily focused on reducing interactions on phase L1. This is because, for most of the time, the demand on L1 is higher than on the other two phases, prompting EnergyR to link PV and home battery resources to this phase. It is important to note that phase switching frequency is constrained by numerous factors; in this study, it is set to 15-minute intervals.

Fig. 5.7 compares the performance of SC-TP, SC-SP, and TC-TP topologies over a 10-day continuous operation period during spring, when PV power generation is at its moderate level. Solar power production and load demand are identical across all phases in each scenario, as shown in Fig. 5.7(a). Additionally, Fig. 5.7(b) compares each scenario's interaction with the electricity grid. The SC-SP topology exhibits the highest energy exchange and, consequently, the lowest self-sufficiency ratio, whereas the TC-TP topology achieves the lowest energy exchange and the highest self-sufficiency ratio. Notably, phase balance is prioritized as the optimization objective across all scenarios rather than maximizing self-sufficiency. This leads to slightly lower self-sufficiency ratios than if the optimization had focused solely on maximizing self-sufficiency.

Additionally, a comparison of the home battery SoC under both single-phase and multi-phase operating conditions is presented in Fig. 5.7(c). The results indicate that when the EnergyR distributes PV-generated power and stored energy across all phases, battery charge and discharge cycles become more frequent, which diminishes battery's longevity by increasing stress on the battery cells. Notably, when the EnergyR and battery operate in single-phase mode, the energy management algorithm has fewer opportunities to maximize the usage of locally generated energy and the self-consumption ratio. This leads to an increased need for PV curtailment or grid injection when the battery is fully charged, ultimately resulting in lower battery utilization. However, this approach helps maintain a higher battery health ratio.

A comparison of various operational modes indicates that the TC-TP topology is the least effective at protecting battery cells from rapid degradation, while the SC-SP mode achieves the highest battery health status during operation compared to other modes. Under high solar energy availability, the performance of the SC-TP mode is comparable to, though slightly better than, the TC-TP mode in terms of battery longevity. However, when solar energy is limited, the SC-TP mode outperforms TC-TP by using the home battery less frequently.

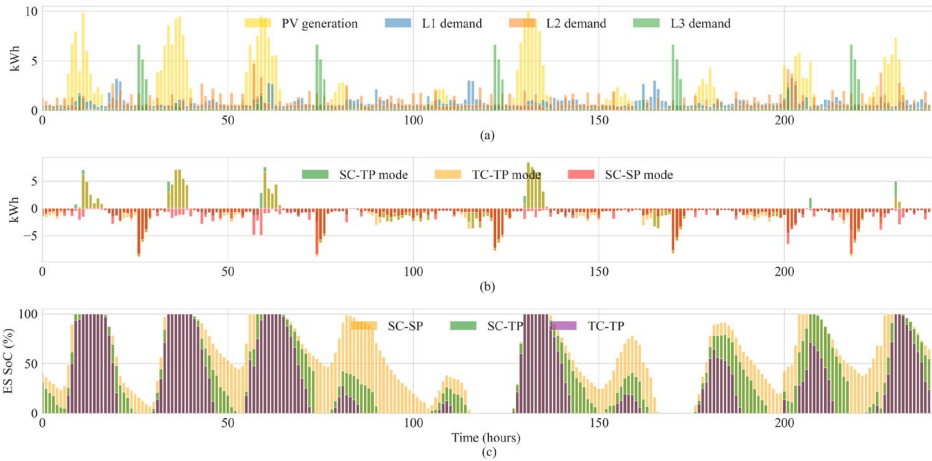


Figure 5.7 Comparison of the three different topologies performances during first 10 days in May-2022. (a) Load demand and PV generation. (b) Amount of energy exchange with the electricity grid (negative values show imported energy and positive values show exported energy to the electricity grid). (c) home battery SoC (%) level during systems' operation.

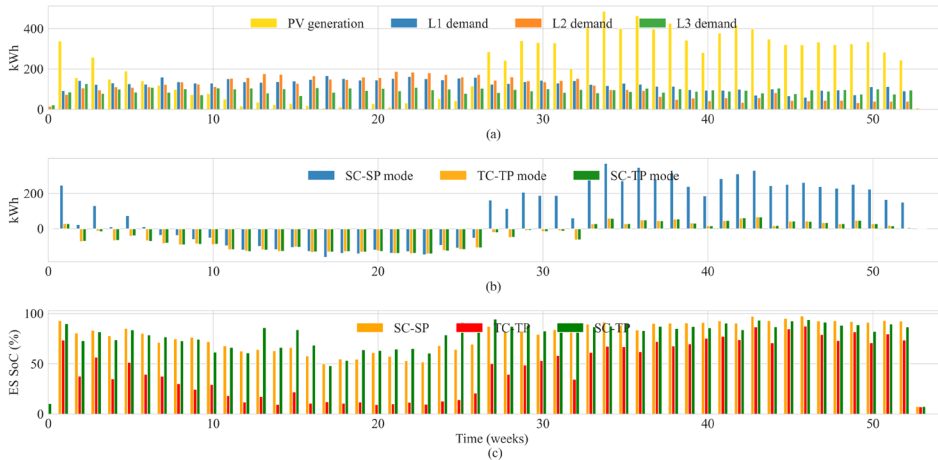


Figure 5.8 Comparison of the topologies based on weekly aggregated performance over a year: (a) Total load demand and PV generation for each week; (b) Total weekly energy exchange with the electricity grid; (c) Average weekly SoC (%) ratio of the home battery.

While this improved longevity is beneficial, it may not fully meet end-user expectations, as reduced battery operation increases reliance on the electricity grid to meet demand. Therefore, it is essential for the energy management unit to account for battery degradation costs, but it should not behave so conservatively and thus lower the system performance. The proper strategy should be able to prioritize resource allocation based on electricity tariffs. Fig. 5.8 presents a comparison of the performance of the investigated operational modes over a year. From a broader perspective, the differences between topologies become clearer. For example, during winter, particularly between weeks

10 and 20, when solar generation is exceptionally low, grid interactions are quite similar across modes. However, during peak solar generation, particularly between weeks 30 and 50, the SC-SP topology shows significantly higher grid interactions, leading to increased grid power flow and potential congestion. This outcome further highlights the limitations of the SC-SP topology in maximizing PV self-consumption, as most of the on-site generated solar power is injected back into the grid. However, two other topologies have been able to utilize locally available renewable sources, to minimize their energy exchanges with the electricity network. The differences between topologies interaction with the electricity network is demonstrated in Fig. 5.8(b).

Fig. 5.8(c) compares the average SoC (%) levels across topologies. It is evident that during periods of low solar generation, TC-TP utilizes home battery more frequently than the other two topologies. However, when solar generation increases, the SoC (%) levels between topologies become more similar, though TC-TP still charges and discharges the batteries more often. Interestingly, the performance of SC-SP and SC-TP topologies is almost identical, with both benefiting from home battery in an equivalent manner. Finally, Fig. 5.9 shows number of times the EnergyR linked renewable resources to each phase in SC-TP topology. Also, Fig. 5.10 presents a comparison of the cost distribution between different EnergyR components in case of SC-TP and TC-TP.

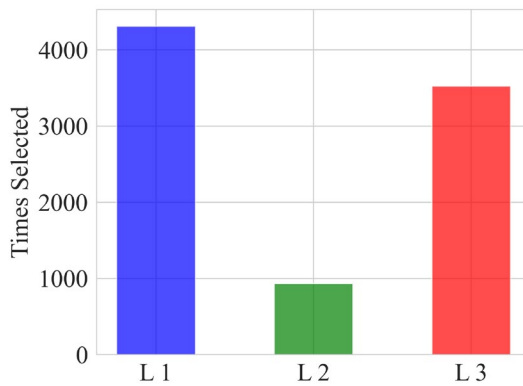


Figure 5.9 A number of times, each phase is connected through EnergyR to the local renewable energy sources.

The cost analysis is based on the retail price of the components used for a single prototype. It includes the cost of all semiconductors and passive components, such as heatsinks and inductors. However, the cost of the printed circuit board and the enclosure used for the demonstrator are not included in the calculations, as these costs are not representative and strongly depend on scaling.

The total cost of the considered components for the prototype of SC-TP EnergyR is approximately 1700 euros. The diagram shows that the most expensive part of the prototype relates to the auxiliary (common) components, such as power supply circuits, heatsinks, and relays. The prototype of the TC-TP EnergyR was not assembled or used in real tests. However, its cost was evaluated based on a bill of materials collected to design the prototype. The total power of the three-phase inverter, as well as the power of the single-phase inverter, was the same, following the same concept with a common ground approach.

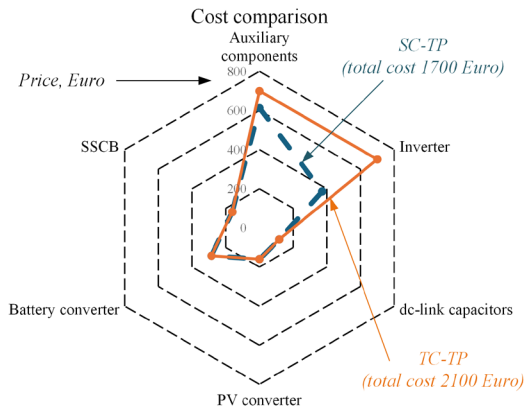


Figure 5.10 Cost comparison diagram which shows the cost distribution between different EnergyR components in case of SC-TP and TC-TP.

The overall cost of the components included in the analysis is around 2100 euros, which is 25% higher than that of the SC-TP energy router. This cost increase is attributed to the higher number of components required for redundant circuits to connect PV setup and home battery to all three phases. It should be noted that the absolute values presented in this thesis cannot be directly used for primary cost estimation, as they strongly depend on scaling, the supply chain during production, and auxiliary circuit optimization. However, the relative comparison is reliable and can be used for cost analysis.

5.3 Summary

This chapter presented a comprehensive approach to optimizing the size and economic feasibility of residential energy systems, focusing on PV systems, battery storage, and Home EMS. A techno-economic model is developed to estimate realistic payback periods by accounting for battery degradation, PV performance, market tariffs, inflation, and user-specific load profiles. The findings show that usage strategies and battery size significantly impact lifespan and return on investment.

Furthermore, simulation results based on real residential data demonstrate that PV size is the dominant factor influencing profitability, while battery size has a secondary role. PV systems above 9 kWp typically offer the shortest payback times, especially under feed-in tariff policies. Systems below a certain size threshold are not economically viable.

To address hardware-related cost and performance trade-offs, a novel SC-TP EnergyR control strategy is introduced. This control algorithm is designed to maximize phase power balances across grid phases and reduces phase imbalances. Compared to other topologies, the SC-TP offers a favorable balance between cost, performance, and battery health, making it a suitable solution for scalable and sustainable residential EMS.

This chapter is based on paper I and paper IV and addresses research task number 2 and hypothesis number 1 and 2. The main contribution is development of an optimal size calculator tool and phase balancing strategy for a SC-TP energy router.

6 Performance Report, and Future Direction

6.1 Performance Report

To validate the proposed Building EMS, a DC-aware hybrid platform was experimentally realized and tested in a nZEB at TalTech campus. As stated before, the system integrates hardware and software components designed to optimize power flows between PV generation, home battery, building loads, and the utility grid while supporting both AC and DC distribution. The overview of the setup is proposed at Introduction section, Fig. 1.2. As evidence in Fig. 5.4, in this platform two DC buses are provided. A 350 V non-isolated bus for heavy DC loads and BESS interfacing. A 48 V isolated bus for light DC loads and consumer devices through smart sockets with USB Type-C outputs. Furthermore, The EnergyR supports three standard communication protocols (CAN, Modbus, Wi-Fi).

In this setup, a Raspberry Pi 4 is used as the HPPU. The HPPU exchanges data with components, including ES, the Graphical UI, EnergyR, load simulator, cloud, and the internet. During system operation, the HPPU, runs an optimization function based on the collected data, and then sends control commands to the EnergyR to determine the appropriate energy source (PV system, BESS, or grid) for meeting the demand. In current setup, CAN interface is utilized for home battery and building EMS data exchange, and Wi-Fi connection is utilized for connecting a cloud-based server to the edge HPPU and LCD display UI. Fig. 6.1 presents the temporal flowchart of the HPPU's operation.

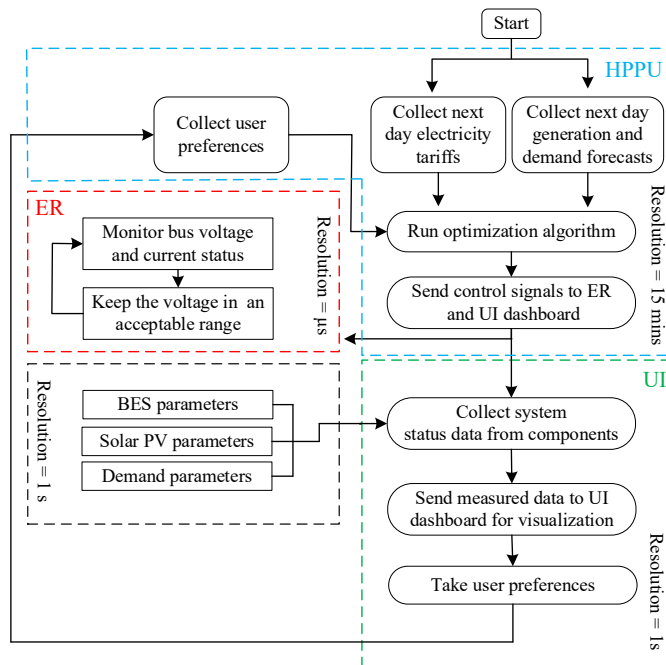


Figure 6.1 Temporal flowchart for HPPU operation and various sections operational resolutions.

As mentioned earlier, since human presence is not feasible in this stage of development, to replicate household demand profiles, a load simulator was developed using an ESP32-based controller which supports integrated radio frequency (RF) communication and Wi-Fi protocol, to activate and deactivate various loads with mechanical relays. Table 6.1 lists the AC and DC loads along with their average daily power consumption. Fig. 6.2 shows the overview of load profile replication with the developed load simulator. Loads connected during testing included typical household appliances, categorized into AC and DC devices. On average, DC-compatible loads accounted for 15% of daily demand, including lighting, personal computers, and small electronics.

Table 6.1 DC and AC loads daily consumption shares.

Name	AC/DC	Avg. demand (kWh/day)	Avg. active time per day*
Lights	DC	0.9	6 hours
Personal computers	DC	1.2	4 hours
Mobile chargers	DC	0.015	3 hours
Vacuum robot	DC	0.1	1 hour
Oven	AC	0.5	10 minutes
Fridge	AC	1.2	24 hours
TV	AC	0.55	5 hours
Dish washer	AC	4.5	2.5 hours
Luandry machine	AC	0.8	0.5 hour
Cook-top stove	AC	6.4	2 hours
Microwave	AC	0.1	10 minutes
Iron	AC	0.17	5 minutes
Hair dryer	AC	0.4	30 minutes
Misselenous	AC	1	1 hour

* Numbers are daily averaged based on total weekly consumption of devices. The house is considered to have a 3-phase connection with 17.32 kW capacity.

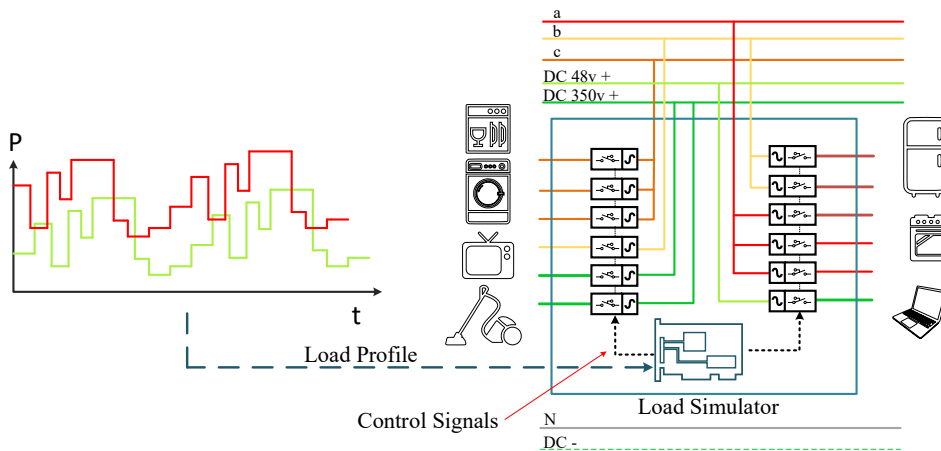


Figure 6.2 Load simulator overview.

6.1.1 Day-Ahead Solar Forecasting Performance

Fig. 6.3 shows a performance of day-ahead solar forecasting unit after deployment in the cloud server. As it is visible, the forecasting tool has in general shows a good performance during the operation. However, since the prediction of cloud movements are not technically possible, so, it is not possible to accurately follow all up and downs of solar power generation during the day. In practice, instead of prediction instantaneous value of solar generation, for residential applications, the precise prediction of accumulated hourly and daily solar generation is considered as a suitable performance. These forecasts are utilized as an input for EMS for energy optimization and resource allocation.

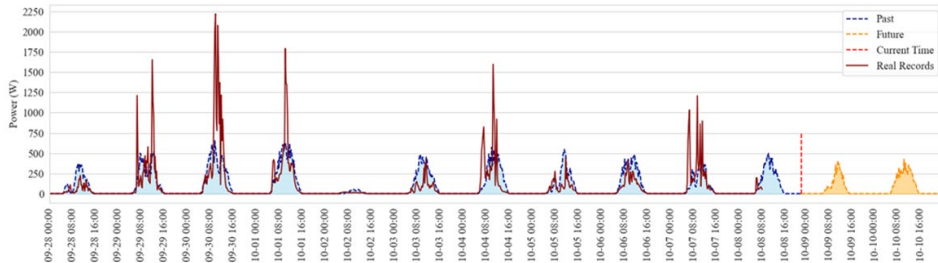


Figure 6.3 Comparison of actual and day ahead forecasting for solar power generation.

6.1.2 Optimization Performance Monitoring

The EMS utilizes hourly electricity tariffs, solar generation forecasts, and a demand profile for optimal resource allocation. In each iteration, it minimizes the objective function based on current and future system states. However, it is important to note that high-level control signals lack the resolution needed to effectively regulate AC and DC bus voltages and prevent any violations related to power quality. Therefore, high-resolution underlying control algorithm is needed to operate at a microsecond scale and ensure uninterrupted power delivery to the end user. EnergyR is using proportional–integral–derivative (PID) controller for this purpose. For example, suppose the EMS sends a command to use solar PV-generated power to meet the demand for the next hour. However, a partially cloudy sky may frequently interrupt solar power generation, making it infeasible to supply the required power at certain moments. In such scenarios, the EnergyR must take over and utilize stored energy or draw electricity from the grid to compensate for the power deficiency.

Fig. 6.4 illustrates the EMS performance over a single day of operation. As shown, the algorithm decides to charge the home Battery when there is a surplus of solar PV power generation and later uses it to supply the demand when electricity tariffs are high. A key highlight is that the EMS chooses to charge the batteries during the early hours of the day. Based on the forecast for solar PV generation on the following day, the algorithm anticipates insufficient renewable power to meet all the demand, and it is evident in Fig. 6.4 (b) that the home battery is discharged during the peak hours at the end of the day. Fig. 6.5 shows the snapshots, collected from power waveforms during EnergyR’s operation including power conversion stages for interfacing various energy sources to the AC and DC loads.

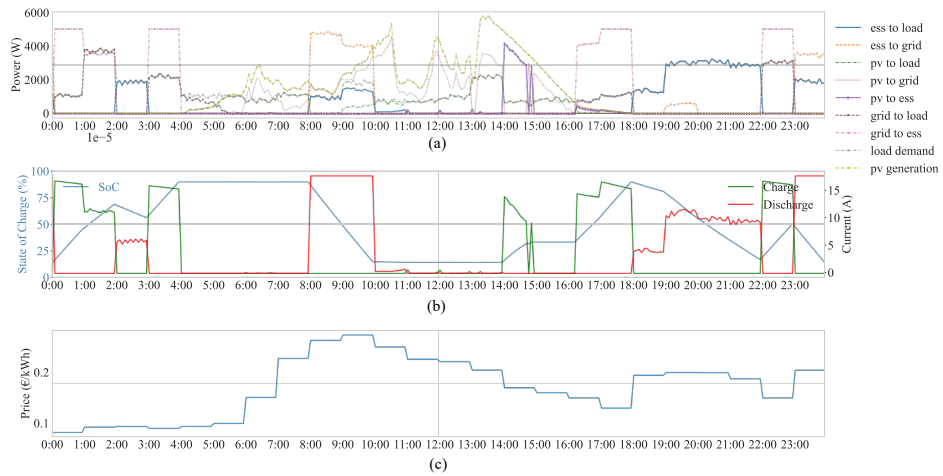


Figure 6.4 EMS performance logs during its operation. (a). Energy flow inside the building electricity network and building-grid power exchanges. (b).

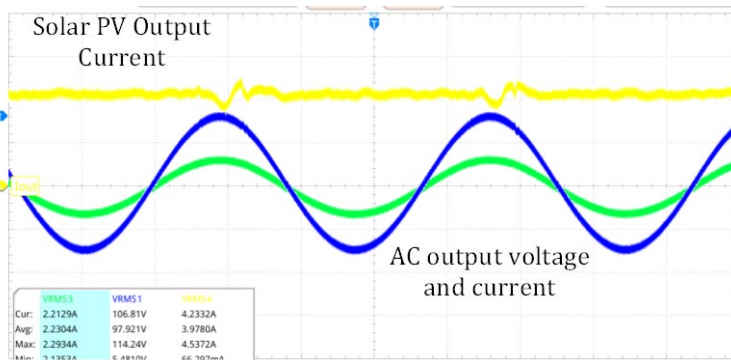


Figure 6.5 EnergyR experimental results supplying AC bus by means of solar power as the main energy source.

From the EMS perspective, both DC and AC loads behave similarly; the primary difference lies in the energy losses incurred when supplying these loads from an AC or DC bus. For example, by connecting devices such as lights, personal computers, mobile chargers, and a vacuum robot to a DC bus, the calculations show a daily energy saving of 110.75 Wh is achievable. This saving is primarily due to the elimination of dual DC/AC/DC conversions, which are independent of EMS performance. Theoretically, the energy saved could reach up to 891.75 Wh per day, representing approximately 7% of the total daily energy consumption.

During wintertime, when solar power is often unavailable due to snow coverage and limited daylight hours, the impact of using DC technology becomes negligible. This is because the loads are primarily supplied by electricity from the AC grid, and an AC/DC conversion stage is still required to feed DC loads. In such periods, economic savings are the focus, achieved through energy arbitrage.

6.1.3 Home Battery Operation Logs

Fig. 6.6 presents the home battery status logs during its charge and discharge cycles. The SoH ratio is also monitored during home battery operation. Capacity fading or degradation is a major factor affecting the economic performance of battery storage systems. Intense battery usage can lead to accelerated degradation; however, batteries often recover a portion of lost capacity after adequate rest periods. In our experiments, based on short-term degradation monitoring and assuming a linear degradation pattern, it is estimated that the home battery will reach 80% of its nominal capacity in approximately 10 years, which is commonly considered the end-of-life threshold. After this point, battery degradation accelerates, and their round-trip efficiency drops.

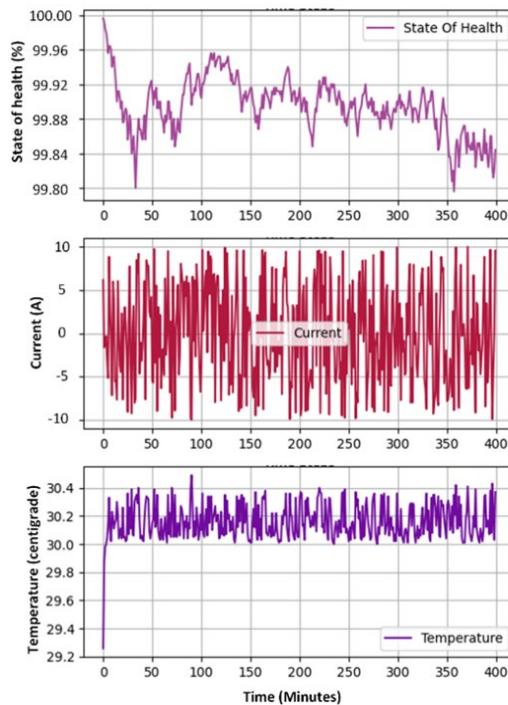


Figure 6.6 Home batter operation logs during charge and discharge cycles.

6.2 Future Direction

6.2.1 Digital Twinning

While the experimental realization of the building EMS has demonstrated its technical feasibility and energy-saving potential, the long-term development of such systems requires deeper integration with Digital Twin (DT) technologies. Digital Twins extend the physical testbed by creating a synchronized virtual replica of the building and its energy infrastructure, enabling advanced monitoring, prediction, optimization, and resilience analysis. Fig. 6.7 represents and compares the DT and physical twin of building EMS. One example of a DT is Google Maps, which models geographical details for services like route planning by synchronizing with real-time data.

DT technology in power systems, could accelerate and streamline the energy transition. Considering buildings as an example, they should autonomously manage their energy tasks, exchange data, and participate in energy markets. They could collaborate with neighbours to form district level energy communities, provide EV charging services, engage in peer-to-peer energy exchange, and rent out infrastructure such as Battery.

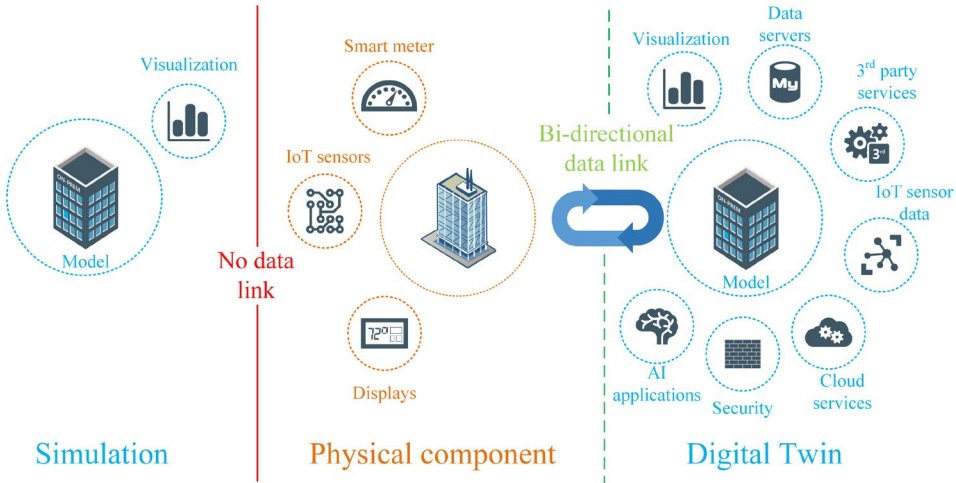


Figure 6.7 Differences between digital twins and simulation models.

The motivation for adopting DT technology stems from several challenges observed during experimental realization:

1. Scalability – Physical testbeds cannot easily replicate diverse demand profiles, weather conditions, or tariff schemes. DTs allow exploration of multiple scenarios in a risk-free virtual environment.
2. Resilience and Reliability – DTs can predict the impact of faults (e.g., inverter malfunction, battery degradation) and suggest corrective strategies before disruptions occur.
3. Market Integration – DTs enable buildings to simulate their participation in energy markets, peer-to-peer trading, and demand response, supporting proactive decision-making.
4. Lifecycle Optimization – By tracking degradation, usage patterns, and efficiency losses, DTs support predictive maintenance and long-term investment planning.

The next phase of this project envisions adopting a hierarchical DT platform. This structure is illustrated in Fig. 6.8. In this approach, a core, component-agnostic DT provides essential functionalities such as data collection, synchronization, and visualization. On top of this foundation, component-specific DTs are developed for PV arrays, batteries, and energy routers. These component DTs can then be aggregated into a building-level DT, which itself can be integrated into larger ecosystems, such as community microgrids or national grid platforms.

This approach simplifies DT development while addressing compatibility challenges. The proposed core DT model is component-agnostic, providing only the essential functionalities required to build component-specific DTs on top of it, much like how a computer Operating System (OS) offers fundamental services while supporting the

installation of application-specific software. For instance, a BESS DT would incorporate a state-of-health estimation module using electrochemical models (e.g., PyBaMM), while a PV DT would integrate irradiance-based generation prediction models (e.g., pvlb). Together, these sub-twins form a comprehensive virtual building energy system capable of high-fidelity simulations, real-time performance tracking, and optimization under diverse operating conditions.

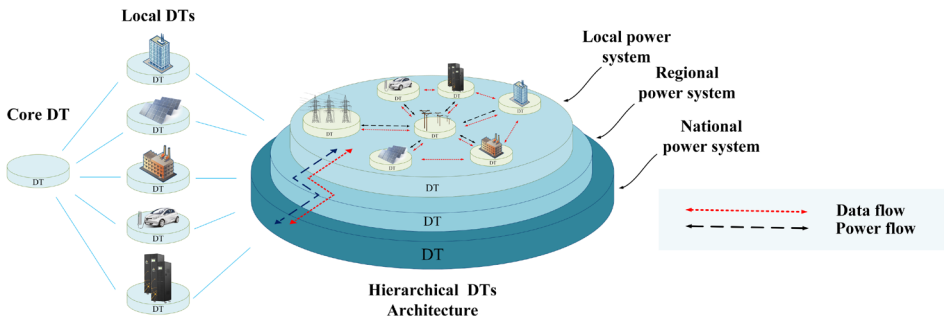


Figure 6.8 Hierarchical platform for DTs and establishing a complex DTs using local DT agents.

6.2.1.1 Core DT

The backbone of proposed layered DT structure is called a core DT. Core DT automates data collection, cleaning, processing, visualization, and finally provides near-real-time knowledge for higher level systems. Figure 3 illustrates the core DT which comprise only base requirements, including: “Data interface” block which facilitates data exchange operations, “DT OS” block which manages DT units’ operations. Fig 6.9 shows various blocks and parts of Core DT.

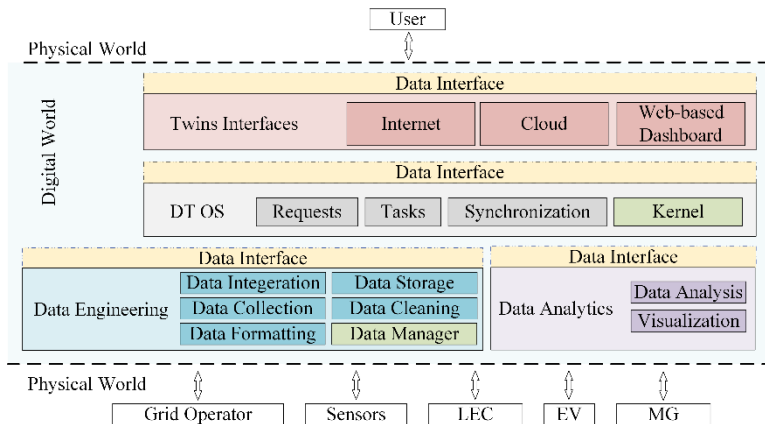


Figure 6.9 Core DT architecture and functional blocks. (LEC: local energy community, MG: micro grids).

Core DT consist of several layers and blocks, such as data interface, DT OS, Data Engineering, Data Analysis, and Twins Interface.

Data Interface: This layer acts as the communication bus and ensures that data exchange among internal DT blocks, between different DTs, and with the physical world adheres to the defined protocols, determined by the origin and destination of data

packets. It provides a framework for message passing across the entire involved parties. All messages—whether sensor readings, control signals, or simulation outputs—must pass through this layer to be routed to the appropriate components.

Users Interface: This layer provides all necessary services for visualizing and interacting with the DT model. Interactions typically occur via mobile applications or web-based dashboards. Users may include general end-users—who adjust system settings, define optimization objectives, or stream reports on techno-economic performance—and experts, who monitor system conditions and manage maintenance activities.

DTOS: This layer acts as the central coordinator for DT operations. The “Kernel” serves as the core execution engine, managing resources and orchestrating inter-component communications. The “Synchronization” block ensures that the DT is consistently updated in real time by aligning data streams and events from physical sensors with their virtual models, thereby maintaining state consistency. The “Task” block allocates process, service, or data as required, executing tasks based on incoming requests while verifying their authorization and legitimacy. Finally, the “Request” block receives, organizes, and processes requests from both external and internal sources, ensuring that all requests are handled efficiently and securely.

Data Engineering: This block is responsible for all data acquisition, processing, formatting, cleaning, and storage. These processes are supervised and coordinated by the “Data Manager” block, which ensures data quality, consistency, and readiness for downstream tasks. Since DTs interact with diverse data types—such as time-series data from sensors, graph-based data representing relationships or network topologies, structured records from enterprise systems, and temporary state data used for real-time decision-making—the choice of database technologies must account for scalability, efficiency, and low-latency performance.

Data Analytics: This layer provides all services required for data analytics, including “visualization”, “forecasting”, “optimization”, etc. Since functionalities in this section are highly dependent on the specific use case for which the DT is designed, as well as the physical asset it represents, more details will be provided in the applied case study.

6.2.1.2 nZEB DT

A nZEB DT can be created by adding building-specific functionalities to the core DT, as illustrated in Fig. 6.10. The modular framework allows for the integration of component-specific features. For example, the “High-Fidelity Simulation” section supports the modeling, simulation, and power flow calculations of the building’s electricity network. The “Simulation Models Manager” ensures that the digital model consistently reflects the building’s electricity network, including loads and energy sources. “Component Models” provide services for holding and delivering component-specific models, such as loads and home battery models. This framework, also, introduces advanced DT functionalities to buildings, such as “building as a service”.

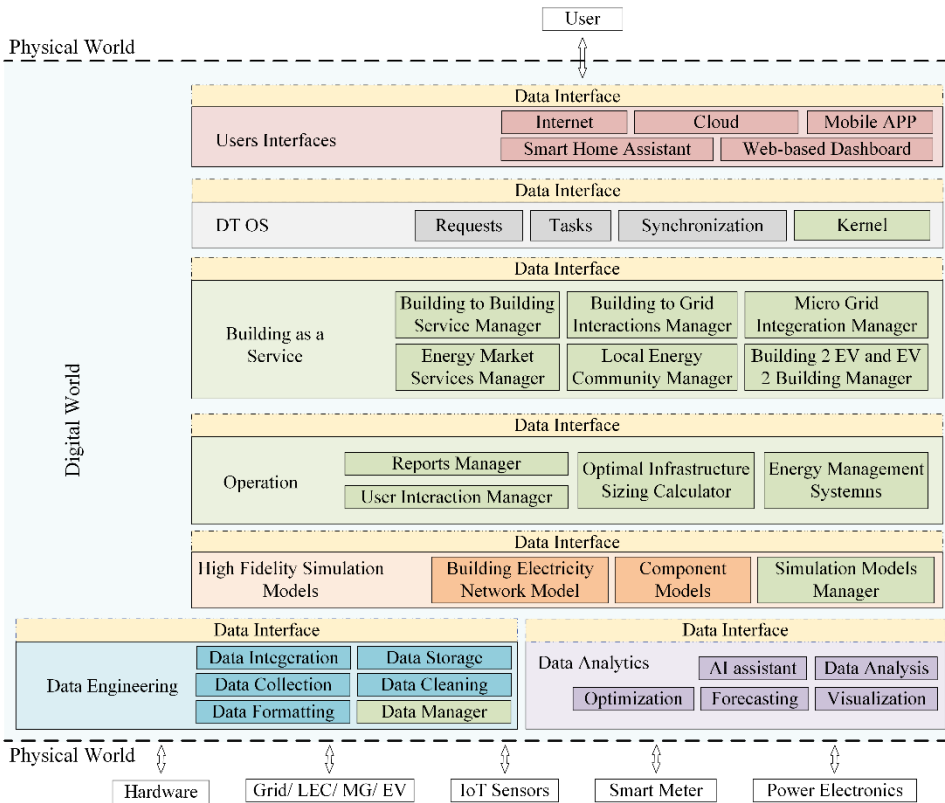


Figure 6.10 Building DT architecture developed based on core DT framework. (IoT: internet of things).

Integrating DT technologies into building EMS platforms, could potentially result in several benefits. However, this integration also may rise several challenges as a side effect, enabling virtual commissioning, enabling synergetic operation of building as an energy block in future intelligent power networks, and introducing better user experience and interactions with a building EMS. Furthermore, several challenges of this technology are, lack of standardization, and universal design, data exchange latency, higher computational complexity, and security issues.

6.3 Summary

This chapter presented the performance evaluation of the proposed DC-aware EMS and outlined future directions with a focus on DT integration. The performance report highlighted the successful experimental realization of the hybrid AC/DC building EMS. The experiments demonstrated the effectiveness of the platform in coordinating renewable generation, storage, and grid interaction. However, since integration of several layers of control, condition monitoring, and optimization is necessary for a safe and reliable operation of building EMS platforms, thus it is necessary to extensively test such systems before real deployment of them for normal user usage.

Looking forward, the chapter highlighted DT integration as the next step in building EMS platforms. A DT enables the creation of a synchronized virtual replica of the building and its energy infrastructure, providing advanced monitoring, prediction, optimization, and resilience analysis. The envisioned hierarchical DT framework builds upon a core

model that offers essential functionalities such as data synchronization, engineering, and analytics. This approach facilitates scalability, predictive maintenance, market participation, and improves user interaction. Nevertheless, the integration of DT technology presents challenges related to standardization, data exchange latency, computational requirements, and cybersecurity, which must be addressed in future research.

This chapter is based on [43] and [59] and addresses research task number 5. The main contribution is experimentally testing the proposed system in a controlled environment and ideal conditions.

7 Conclusions

Integration of solar power harvesting facilities is an initial step toward making buildings energy independent. The next stage in boosting building energy autonomy is the installation of local energy storage systems, such as home batteries. In large-scale buildings, such as shopping malls and complexes, hydrogen storage technology can also be a viable option. Having both generation and storage facilities may appear to be the ultimate solution for maximizing building self-sufficiency. However, this thesis demonstrates that without an intelligent supervisory EMS, it is highly challenging to maintain synergy between energy generation and consumption.

Moreover, a high-performance EMS requires accurate estimation of both the available solar energy and the expected demand in advance to solve the optimization problem and define an optimal (or suboptimal) trajectory and energy usage scenario. Insights into near-future possibilities can be obtained either through classical forecasting methods or by applying data-driven AI- and ML-based tools. Thanks to advancements in computational power and the abundance of data, AI and ML models often outperform classical methods. However, in data-scarce scenarios, such as newly installed solar power generation systems, AI models exhibit poor performance. To address this challenge, physics-based models can be coupled with AI models to improve accuracy when insufficient historical data is available.

Because of the stochastic nature of both energy generation and consumption, it is impossible to predict future trends with perfect accuracy. Thus, relying solely on deterministic approaches for scenario generation and power flow optimization reduces EMS performance. A better approach is to integrate system randomness and employ stochastic optimization methods, enabling the system to handle uncertainties and generate solutions that may not be optimal for a single scenario but are optimal across multiple possible scenarios.

Power electronics also play a crucial role in improving the energy performance of buildings. Based on the investigations in this thesis, the author argues that integrating DC solutions is essential for future building technologies. Utilizing a DC network inside a building offers several benefits, including reduced energy loss from power conversions and the elimination of AC/DC conversion units inside electronic devices. Furthermore, with the emergence of DC microgrids and DC neighbourhoods, integrating a DC-equipped building into larger systems becomes more straightforward compared to AC-based buildings.

Finally, as IoT technology becomes an integral part of modern buildings, which are equipped with air conditioning systems, smart devices, and advanced sensors, the digital twinning of building from energy perspective can further enhance performance of EMSs. This approach facilitates the transition toward fully autonomous building energy management at an advanced level. Here, “advanced” refers to a stage in which a building can handle all energy-related tasks without user intervention, achieve defined goals (e.g., minimum energy bills, maximum energy autonomy), and autonomously interact with intelligent power systems, other buildings, EVs, e-scooters, etc. Potential applications include participation in energy markets, providing utility services, and enabling peer-to-peer energy trading.

As a results of thesis, author can claim the following:

- Stochastic optimization techniques outperform deterministic and rule-based solutions; however, any high-level optimization algorithm is only effective

when supported by a high-resolution low-level control algorithm that can monitor and react to system status changes at microsecond resolution.

- Physics-based models can enhance the performance of data-driven and ML-based models when the amount of historical data is insufficient for training. Furthermore, physics-based models can be used for synthetic data generation to train ML models, which helps improve their performance. Based on studied data and PV setup, including physics-based model, ultimately improves a ML models accuracy by 20%.
- A hybrid AC/DC solution for residential buildings' internal electricity distribution networks improves energy performance only if sufficient solar energy is available. In winter, particularly in northern parts of Europe where solar irradiance is very limited, a hybrid solution may decrease a building's energy performance.
- For residential applications in Estonia, considering climate conditions and electricity prices based on data from 2022–2024, not all combinations of PV–home battery setups are beneficial. Case studies show that PV systems smaller than 4 kWp and home batteries below 5 kWh are not profitable once degradation effects are considered.

Future work will focus on improving EMS performance through benchmarking of novel stochastic and data-driven (both model-free and model-based) control methodologies, as well as extended field testing to monitor the seasonal effects on system performance.

List of Figures

Figure 1.1 California’s duck curve trend with higher penetration of distributed renewable energy resources. CAISO lowest net load day each spring (March–May, 2015–2023) [19].	15
Figure 1.2 Experimental setup.....	18
Figure 2.1 Some samples of commercialized solutions for home EMS from ENPHASE and LG ELECTRONICS companies.....	21
Figure 2.2 Proposed building DC-aware EMS ecosystem.....	23
Figure 2.3 Proposed system components from data flow perspective.....	24
Figure 2.4 (a) Web based EMS user interface and dashboard, (b)Local GUI running on edge device for user interaction with a system.....	25
Figure 2.5 Designed user interface panel, navigation bar, and different sections for user interaction and system status reporting.....	25
Figure 2.6 Proposed six-layer software architecture and each layers dedicated responsibilities.....	26
Figure 2.7 EnergyR internal topology and TRL-5 prototype.....	27
Figure 2.8 Demand simulation controller board.....	27
Figure 2.9 BMS configuration and connectivity, and home battery assembly and test stages in the laboratory.....	28
Figure 3.1 Proposed framework for day-ahead solar generation forecasting.....	29
Figure 3.2 PV module layout and equivalent circuit model of solar cells connected in the series format.....	30
Figure 3.3 Solar radiation predictions for the same day and location, collected from different weather service providers (a). Predicted temperatures for the same day and location collected from different weather service providers (b).....	31
Figure 3.4 The detailed pipeline of the proposed hybrid ML model.....	32
Figure 3.5 Proposed SAED based solar power forecasting model.....	32
Figure 3.6 Sliding window technique and SAED network input data structure.....	33
Figure 3.7 Proposed forecasting framework comparison with measured data.....	35
Figure 3.8 Comparison between performance of the hybrid ML-PBM model with PBM model. Hybrid Model (a), PBM model (b).....	35
Figure 3.9 Hourly averaged electricity demand forecasting using LSTM network.....	36
Figure 4.1 Home EMS system overview, considering integration of renewable energies, battery energy storage systems, and various loads.....	38
Figure 4.2 Deterministic Energy flow Optimization flowchart.....	41
Figure 4.3 A flowchart of proposed stochastic optimization method based on forecast driven and historical data scenario generation.....	45
Figure 4.4 Abstract representation of obtaining \mathcal{GPH} from historical solar PV power generation records. Solar irradiance curves are extracted from estimation of the hourly global solar irradiation based on numerical weather predictions.....	46
Figure 4.5 Daily scenario generation based on calculated probability matrices and random candidate selection.....	47
Figure 4.6 Energy bill, Self consumption, and energy autonomy ration correlations, applying different scenarios.....	51
Figure 4.7 Error injections to PV and load profiles. (a) PV profile fluctuations with error injections. (b) Load profile fluctuations with error injections.....	53

Figure 4.8 System variables deviations from ideal condition in presence of error injection to PV and Load profiles. (a) ESS SoC level deviations from ideal condition. (b) PV to load delivered power deviations from ideal condition. (c) Grid to load delivered power deviations from ideal condition. (d) ESS to load delivered power deviations from ideal condition.....	54
Figure 5.1 Battery degradation ratio for all possible combinations of PV and home battery. (a) Degradation level dependency on BESS size. (b) Degradation level dependency on solar generation setup size.....	61
Figure 5.2 Economic performance analysis of the case study house in Estonia.....	61
Figure 5.3 Discounted payback calculation of PV-home battery combinations. (a) For each BESS size when the PV size is swept in its range. (b) For each PV size when the BESS size is swept in its range. The Y axis has a logarithmic scale in (b).....	62
Figure 5.4 Abstract view of the proposed single cell topology connected to three-phase terminal with EV charger, PV, and ES integration and its experimental realization [58]....	64
Figure 5.5 Imbalanced and balanced phase power range comparison assuming ideal phase balancing possibilities.....	65
Figure 5.6 Normalized radar chart comparison for all topologies.....	67
Figure 5.7 Comparison of the three different topologies performances during first 10 days in May-2022. (a) Load demand and PV generation. (b) Amount of energy exchange with the electricity grid (negative values show imported energy and positive values show exported energy to the electricity grid). (c) home battery SoC (%) level during systems' operation.....	69
Figure 5.8 Comparison of the topologies based on weekly aggregated performance over a year: (a) Total load demand and PV generation for each week; (b) Total weekly energy exchange with the electricity grid; (c) Average weekly SoC (%) ratio of the home battery.....	69
Figure 5.9 A number of times, each phase is connected through EnergyR to the local renewable energy sources.....	70
Figure 5.10 Cost comparison diagram which shows the cost distribution between different EnergyR components in case of SC-TP and TC-TP.....	71
Figure 6.1 Temporal flowchart for HPPU operation and various sections operational resolutions.....	72
Figure 6.2 Load simulator overview.....	73
Figure 6.3 Comparison of actual and day ahead forecasting for solar power generation..	74
Figure 6.4 EMS performance logs during its operation. (a). Energy flow inside the building electricity network and building-grid power exchanges. (b).....	75
Figure 6.5 EnergyR experimental results supplying AC bus by means of solar power as the main energy source.....	75
Figure 6.6 Home batter operation logs during charge and discharge cycles.....	76
Figure 6.7 Differences between digital twins and simulation models.....	77
Figure 6.8 Hierarchical platform for DTs and establishing a complex DTs using local DT agents.....	78
Figure 6.9 Core DT architecture and functional blocks. (LEC: local energy community, MG: micro grids).....	78
Figure 6.10 Building DT architecture developed based on core DT framework. (IoT: internet of things).....	80

List of Tables

Table 2.1 Home EMS solutions comparison.....	20
Table 4.1 Performance comparison between various EMS scenarios	52
Table 5.1 Circular capacity fading model's constant parameters	58
Table 5.2 Financial and technical parameters.....	58
Table 5.3 Most profitable combinations' economical results.....	63
Table 5.4 EnergyR technical parameters.....	65
Table 6.1 DC and AC loads daily consumption shares.....	73

References

- [1] H. N. Hokmabad, O. Husev, J. Kurnitski, and J. Belikov, 'Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications', *Sustainable Energy, Grids and Networks*, vol. 38, p. 101385, 2024.
- [2] S. Gong, V. Čuk, and J. F. G. Cobben, 'Multidimensional Hosting Capacity Region in Dutch MV Grid Congestion Management', *IEEE Transactions on Power Systems*, vol. 39, no. 3, pp. 5024–5035, 2024.
- [3] H. Zhou, L. Lu, M. Wei, L. Shen, and Y. Liu, 'Robust Scheduling of a Hybrid Hydro/Photovoltaic/Pumped-Storage System for Multiple Grids Peak-Shaving and Congestion Management', *IEEE Access*, vol. 12, pp. 22230–22242, 2024.
- [4] J. Wang, M. Liu, and H. Wu, 'The Demand-Side Management and Control of Smart Grids Based on Weighted Network Congestion Games', *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 43–52, 2025.
- [5] J. Wang, M. Liu, and H. Wu, 'The Demand-Side Management and Control of Smart Grids Based on Weighted Network Congestion Games', *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 43–52, 2025.
- [6] W. Velasquez, G. Z. Moreira-Moreira, and M. S. Alvarez-Alvarado, 'Smart Grids Empowered by Software-Defined Network: A Comprehensive Review of Advancements and Challenges', *IEEE Access*, vol. 12, pp. 63400–63416, 2024.
- [7] K. Moslehi and R. Kumar, 'A Reliability Perspective of the Smart Grid', *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 57–64, 2010.
- [8] X. Cao, X. Dai, and J. Liu, 'Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade', *Energy and Buildings*, vol. 128, pp. 198–213, 2016.
- [9] M. Santamouris and K. Vasilakopoulou, 'Present and future energy consumption of buildings: Challenges and opportunities towards decarbonisation', *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 1, p. 100002, 2021.
- [10] T. Harputlugil and P. de Wilde, 'The interaction between humans and buildings for energy efficiency: A critical review', *Energy Research & Social Science*, vol. 71, p. 101828, 2021.
- [11] M. Elkazaz, M. Sumner, E. Naghiyev, Z. Hua, and D. W. P. Thomas, 'Techno-Economic Sizing of a community battery to provide community energy billing and additional ancillary services', *Sustainable Energy, Grids and Networks*, vol. 26, p. 100439, 2021.
- [12] Garbesi K, Vossos V, Shen H. Catalog of DC appliances and power systems. Tech. rep BNL-5364E, Lawrence Berkeley National Laboratory; 2011.
- [13] I. Calero, C. A. Cañizares, K. Bhattacharya, and R. Baldick, 'Duck-Curve Mitigation in Power Grids With High Penetration of PV Generation', *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 314–329, 2022.
- [14] N. Mahmud and A. Zahedi, 'Review of control strategies for voltage regulation of the smart distribution network with high penetration of renewable distributed generation', *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 582–595, 2016.
- [15] L. R. Visser, E. M. B. Schuurmans, T. A. AlSkaif, H. A. Fidder, A. M. van Voorden, and W. G. J. H. M. van Sark, 'Regulation strategies for mitigating voltage fluctuations induced by photovoltaic solar systems in an urban low voltage grid', *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107695, 2022.

- [16] O. Pereira, M. Parajeles, B. Zúñiga, J. Quirós-Tortós, and G. Valverde, 'Hosting Capacity Estimation for Behind-the-Meter Distributed Generation', *IEEE Transactions on Power Systems*, vol. 39, no. 3, pp. 4911–4923, 2024.
- [17] M. Rylander, J. Smith, and W. Sunderman, 'Streamlined Method for Determining Distribution System Hosting Capacity', *IEEE Transactions on Industry Applications*, vol. 52, no. 1, pp. 105–111, 2016.
- [18] S. Wang, S. Chen, L. Ge, and L. Wu, 'Distributed Generation Hosting Capacity Evaluation for Distribution Systems Considering the Robust Optimal Operation of OLTC and SVC', *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1111–1123, 2016.
- [19] U.S. Energy Information Administration, 'As solar capacity grows, duck curves are getting deeper in California', 2023. [Online]. Available: <https://www.eia.gov/todayinenergy/detail.php?id=56880>. [Accessed: 30-May-2025].
- [20] J. Faraji, F. Vallée, and Z. De Grève, 'A Preference-Informed Energy Sharing Framework for a Renewable Energy Community', *IEEE Transactions on Energy Markets, Policy and Regulation*, vol. 2, no. 4, pp. 503–518, 2024.
- [21] Y. Zahraoui, T. Korötko, A. Rosin, T. E. K. Zidane, H. Agabus, and S. Mekhilef, 'A Competitive Framework for the Participation of Multi-Microgrids in the Community Energy Trading Market: A Case Study', *IEEE Access*, vol. 12, pp. 68232–68248, 2024.
- [22] P. Afzali, A. Rajaei, M. Rashidinejad, and H. Farahmand, 'Peer-to-Peer Energy Trading Among Prosumers in Energy Communities Based on Preferences Considering Holacracy Structure', *IEEE Transactions on Engineering Management*, vol. 71, pp. 7756–7767, 2024.
- [23] R. Li and S. You, 'Exploring potential of energy flexibility in buildings for energy system services', *CSEE Journal of Power and Energy Systems*, vol. 4, no. 4, pp. 434–443, 2018.
- [24] S. Nikkiah, A. Allahham, M. Royapoor, J. W. Bialek, and D. Giaouris, 'Optimising Building-to-Building and Building-for-Grid Services Under Uncertainty: A Robust Rolling Horizon Approach', *IEEE Transactions on Smart Grid*, vol. 13, no. 2, pp. 1453–1467, 2022.
- [25] A. Saif, S. K. Khadem, M. F. Conlon, and B. Norton, 'Impact of Distributed Energy Resources in Smart Homes and Community-Based Electricity Market', *IEEE Transactions on Industry Applications*, vol. 59, no. 1, pp. 59–69, 2023.
- [26] L. Che, M. Shahidehpour, A. Alabdulwahab, and Y. Al-Turki, 'Hierarchical Coordination of a Community Microgrid With AC and DC Microgrids', *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 3042–3051, 2015.
- [27] H. Nagpal, I.-I. Avramidis, F. Capitanescu, and A. G. Madureira, 'Local Energy Communities in Service of Sustainability and Grid Flexibility Provision: Hierarchical Management of Shared Energy Storage', *IEEE Transactions on Sustainable Energy*, vol. 13, no. 3, pp. 1523–1535, 2022.
- [28] W. Guedes, C. Oliveira, T. A. Soares, B. H. Dias, and M. Matos, 'Collective Asset Sharing Mechanisms for PV and BESS in Renewable Energy Communities', *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 607–616, 2024.
- [29] C. Gschwendtner, S. R. Sinsel, and A. Stephan, 'Vehicle-to-X (V2X) implementation: An overview of predominate trial configurations and technical, social and regulatory challenges', *Renewable and Sustainable Energy Reviews*, vol. 145, p. 110977, 2021.

- [30] J. Zupančič, B. Filipič, and M. Gams, 'Genetic-programming-based multi-objective optimization of strategies for home energy-management systems', *Energy*, vol. 203, p. 117769, 2020.
- [31] H. Yang, Z. Gong, Y. Ma, L. Wang, and B. Dong, 'Optimal two-stage dispatch method of household PV-BESS integrated generation system under time-of-use electricity price', *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106244, 2020.
- [32] S. Yang, H. O. Gao, and F. You, 'Building electrification and carbon emissions: Integrated energy management considering the dynamics of the electricity mix and pricing', *Advances in Applied Energy*, vol. 10, p. 100141, 2023.
- [33] S. Sen and M. Kumar, 'Distributed-MPC Type Optimal EMS for Renewables and EVs Based Grid-Connected Building Integrated Microgrid', *IEEE Transactions on Industry Applications*, vol. 60, no. 2, pp. 2390–2408, 2024.
- [34] C. Xia, W. Li, X. Chang, F. C. Delicato, T. Yang, and A. Y. Zomaya, "Edge-based energy management for smart homes," in *Proc. IEEE 16th Intl Conf Dependable, Autonomic Secure Comput., 16th Intl Conf Pervasive Intell. Comput., 4th Intl Conf Big Data Intell. Comput. Cyber Sci. Technol. Congress(DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 849–856.
- [35] P. Lissa, C. Deane, M. Schukat, F. Seri, M. Keane, and E. Barrett, 'Deep reinforcement learning for home energy management system control', *Energy and AI*, vol. 3, p. 100043, 2021.
- [36] M. A. Alghassab, 'Fuzzy-based smart energy management system for residential buildings in Saudi Arabia: A comparative study', *Energy Reports*, vol. 11, pp. 1212–1224, 2024.
- [37] A. Basit, G. A. S. Sidhu, A. Mahmood, and F. Gao, "Efficient and autonomous energy management techniques for the future smart homes," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 917–926, Mar. 2017.
- [38] S. Paul and N. P. Padhy, 'Real-Time Energy Management for Smart Homes', *IEEE Systems Journal*, vol. 15, no. 3, pp. 4177–4188, 2021.
- [39] C. A. Correa-Florez, A. Gerossier, A. Michiorri, and G. Kariniotakis, 'Stochastic operation of home energy management systems including battery cycling', *Applied Energy*, vol. 225, pp. 1205–1218, 2018.
- [40] Z. Wang, F. Xiao, Y. Ran, Y. Li, and Y. Xu, 'Scalable energy management approach of residential hybrid energy system using multi-agent deep reinforcement learning', *Applied Energy*, vol. 367, p. 123414, 2024.
- [41] Y.-H. Lin, H.-S. Tang, T.-Y. Shen, and C.-H. Hsia, 'A Smart Home Energy Management System Utilizing Neurocomputing-Based Time-Series Load Modeling and Forecasting Facilitated by Energy Decomposition for Smart Home Automation', *IEEE Access*, vol. 10, pp. 116747–116765, 2022.
- [42] V. Lo Brano, A. Orioli, G. Ciulla, and A. Di Gangi, 'An improved five-parameter model for photovoltaic modules', *Solar Energy Materials and Solar Cells*, vol. 94, no. 8, pp. 1358–1370, 2010.
- [43] H. N. Hokmabad, O. Husev, Pedro P. Vergara, J. Kurnitski, D. Vinnikov and J. Belikov, 'Building-Level DC-Aware Energy Management System: Experimental Realization and Outcomes', in *2025 IEEE 7TH International Conference on DC Microgrids (ICDCM 2025)*, 2025.

- [44] H. Tian, F. Mancilla-David, K. Ellis, E. Muljadi, and P. Jenkins, 'A cell-to-module-to-array detailed model for photovoltaic panels', *Solar energy*, vol. 86, no. 9, pp. 2695–2706, 2012.
- [45] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, 'A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output', *IEEE transactions on sustainable energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [46] T. Zhang, 'Solving large scale linear prediction problems using stochastic gradient descent algorithms', in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 116.
- [47] M. Nejati and N. Amjady, 'A new solar power prediction method based on feature clustering and hybrid-classification-regression forecasting', *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1188–1198, 2021.
- [48] H. N. Hokmabad, O. Husev, J. Belikov, D. Vinnikov, and E. Petlenkov, 'Energy Storage and Forecasting Error Impact Analysis in Photovoltaic Equipped Residential Nano-Grids', in *2023 IEEE 17th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2023, pp. 1–6.
- [49] U. G. K. Mulleriyawage and W. X. Shen, 'Optimally sizing of battery energy storage capacity by operational optimization of residential PV-Battery systems: An Australian household case study', *Renewable Energy*, vol. 160, pp. 852–864, 2020.
- [50] M. Swierczynski, D.-I. Stroe, A.-I. Stan, R. Teodorescu, and S. K. Kær, 'Lifetime Estimation of the Nanophosphate LiFePO₄ Battery Chemistry Used in Fully Electric Vehicles', *IEEE Transactions on Industry Applications*, vol. 51, no. 4, pp. 3453–3461, 2015.
- [51] L. Lam and P. Bauer, 'Practical capacity fading model for Li-ion battery cells in electric vehicles', *IEEE transactions on power electronics*, vol. 28, no. 12, pp. 5910–5918, 2012.
- [52] D. Gardiner, O. Schmidt, P. Heptonstall, R. Gross, and I. Staffell, 'Quantifying the impact of policy on the investment case for residential electricity storage in the UK', *Journal of Energy Storage*, vol. 27, p. 101140, 2020.
- [53] S. Bandyopadhyay, G.R.C. Mouli, Z. Qin, L.R. Elizondo, P. Bauer, 'Techno-economical model based optimal sizing of PV-battery systems for microgrids', *IEEE Trans. Sustain. Energy*, vol. 11 (3) (2019) 1657–1668.
- [54] B. P. Numbi and S. J. Malinga, 'Optimal energy cost and economic analysis of a residential grid-interactive solar PV system- case of eThekweni municipality in South Africa', *Applied Energy*, vol. 186, pp. 28–45, 2017.
- [55] M. Najafzadeh et al., 'Grid-Forming Operation of Energy-Router Based on Model Predictive Control with Improved Dynamic Performance', *Energies*, vol. 15, no. 11, 2022.
- [56] M. Mehrtash, F. Capitanescu, P. K. Heiselberg, T. Gibon, and A. Bertrand, 'An Enhanced Optimal PV and Battery Sizing Model for Zero Energy Buildings Considering Environmental Impacts', *IEEE Transactions on Industry Applications*, vol. 56, no. 6, pp. 6846–6856, 2020.
- [57] S. N. Motapon, E. Lachance, L.-A. Dessaint, and K. Al-Haddad, 'A Generic Cycle Life Model for Lithium-Ion Batteries Based on Fatigue Theory and Equivalent Cycle Counting', *IEEE Open Journal of the Industrial Electronics Society*, vol. 1, pp. 207–217, 2020.

- [58] S. Rahimpour, "Common-Ground Energy Router Structure with Enhanced Reliability and Protection," Ph.D. dissertation, *Tallinn University of Technology*, 2024.
- [59] H. N. Hokmabad, C. E. S. Lekhel, T. H. Shamsavar, P. P. Vergara, O. Husev, and J. Belikov, 'A Unified Hierarchical Digital Twin Platform for Synergistic Management of Low-voltage Electrical Network Components', in *2025 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2025.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Prof. Dmitry Vinnikov, Assoc. Prof. Oleksandr Husev, and Assoc. Prof. Juri Belikov.

Dear Oleksandr and Juri, your endless support, kindness, understanding, professionalism, and profound knowledge have taught me many lessons, in field of study, my professional career, and as a human being. I am deeply grateful for all the moments we shared during these fast-paced four years of my PhD journey at Tallinn University of Technology. I am sincerely thankful to Prof. Vinnikov, Head of the Power Electronics Group at Tallinn University of Technology, for providing continuous support and creating an environment that enabled all members of the group to focus on their research.

During this period, I had the privilege of working and learning alongside many brilliant researchers. I would like to thank Prof. Eduard Petlenkov from the Department of Computer Systems at TalTech for his continuous support and guidance. My heartfelt thanks also go to my colleagues, Oleksandr, Marika, Aleksei, Komeil, Christina, and many others, who helped, collaborated, and shared their knowledge with me.

I gratefully acknowledge the financial support of the Estonian Research Council (grants PRG675, PRG1463, EAG234) and the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230), funded by the Estonian Ministry of Education and Research. I also wish to express my appreciation to the Erasmus program, which supported my research visit to Delft University of Technology. I am sincerely thankful to Prof. Peter Palensky and Assist. Prof. P. P. Vergara for welcoming me to the IEPG group at TU Delft.

I am profoundly grateful to my parents, Hassan and Nasrin, for their unwavering love, support, and encouragement. I am always proud of you, and I hope to make you proud of me in return.

To my wife, best friend, and soulmate, Tala, your both technical and mental helps, patience, support, and constant encouragement have been an endless source of energy throughout this journey. I am deeply thankful to have you by my side.

And finally, to Adrian, my beautiful son. While writing the sixth chapter of this thesis, you came into our lives. Words cannot describe the joy and happiness you have brought us. I dedicate this thesis to you and your mother, Tala. Love you both.

I end with my sincere wish: may peace, happiness, and kindness prevail across the world, forever, for every human.

“Until death, all defeat is psychological.”

Abstract

Energy management system for single-cell three phase energy router in residential applications

This dissertation develops and validates an integrated hardware–software platform for residential Building EMS that couples rooftop PV, a home battery, and flexible loads through a cost-effective, single-cell, three-phase EnergyR. The work is motivated by the growing complexity of low-voltage networks, where behind-the-meter renewables and EV charging intensify variability and phase imbalance; buildings therefore need EMSs that can forecast, optimize, and coordinate power flows while remaining economically viable. The thesis formulates clear aims and hypotheses: optimal sizing can shorten payback times; a single-cell interface can serve three-phase grids; stochastic optimization can outperform deterministic methods under uncertainty; and hybrid physics–data-driven forecasting can improve day-ahead accuracy when historical data are limited.

Methodologically, the platform comprises (i) a hybrid day-ahead PV forecast that fuses a physics-based model with a ML branch (LightGBM regressors conditioned on SYNOP weather codes) and a sequence model (BiLSTM self-attention encoder–decoder), to produce the final prediction; (ii) residential demand forecasting using LSTM/XGBoost; and (iii) an EMS optimizer implemented as heuristic rules, a deterministic MILP in Pyomo/GLPK, and a scenario-based stochastic MILP to explicitly handle forecast and tariff uncertainty.

Across 100 evaluation days, the hybrid forecast outperforms LSTM, BiLSTM, XGBoost, LightGBM, encoder–decoder, physics-only, and persistence baselines, supporting the hypothesis that hybridization improves accuracy in data-scarce, weather-sensitive settings.

A techno-economic framework integrates market tariffs, inflation, PV performance, and an explicit battery degradation model to compute realistic payback times. Case studies with real residential data show PV size dominates profitability while battery capacity plays a secondary role; PV systems > 9 kWp typically yield the shortest payback, whereas undersized systems may be uneconomic.

On the hardware side, the thesis proposes and justifies a SC-TP EnergyR that connects a single conversion cell to all three phases non-simultaneously via controlled relays, enabling phase-aware power dispatch and balancing within the home and at the grid interface. This topology reduces phase imbalance by steering PV/ battery power to the most loaded phase while avoiding two additional conversion cells.

The complete platform is experimentally validated at TalTech’s nZEB. Collectively, the contributions include a cold-start-robust solar forecasting method, a scenario-based stochastic EMS optimizer, an infrastructure sizing tool accounting for battery fade, and the SC-TP EnergyR concept that lowers capex while improving technical performance. These achievements, advance the readiness of residential buildings to participate in future intelligent, renewable-dominant power systems.

Lühikokkuvõte

Energiahaldussüsteem ühelemendilisele kolmefaasilisele energiaruuterile elamutes

Käesolevas väitekirjas arendatakse ja valideeritakse integreeritud riist- ja tarkvaraplatvorm elamute energiahaldussüsteemidele (EHS), mis ühendavad päikesepaneelid, akusalvesti ja paindlikud koormused kulutõhusa ühelemendilise kolmefaasilise energiaruuteri (EnergyR) abil. Töö on motiveeritud madalpingevõrkude kasvavast keerukusest tänu lisanduvatele taastuvenergiaallikatele ja elektriautolaadijatele, mis suurendavad muutlikkust ja faaside tasakaalustamatust. Seetõttu vajavad hooned EHS-e, mis suudavad prognoosida, optimeerida ja koordineerida energiavooge, jäädes samas majanduslikult tasuvaks. Töö sõnastab selged eesmärgid ja hüpoteesid: optimaalne akusuurus võib lühendada tasuvusaega; ühelemendiline võrguliides võib teenindada kolmefaasilisi võrke; stohhastiline optimeerimine võib ebakindluse korral ületada deterministlikke meetodeid; hübriidsete andmepõhiste mudelitega prognoosimine võib parandada ennustuse täpsust, kui ajaloolised andmed on piiratud.

Platvorm koosneb järgmistest komponentidest: (i) hübriidsest päev-ette päikesepaneelide tootlikkuse prognoosist, mis ühendab füüsilise mudeli masinõppe (ML) haruga (LightGBM regresioonid, mis sõltuvad SYNOP ilmaprognoosi koodidest) ja järjestusmudeliga (ingl. BiLSTM self-attention encoder–decoder), mis tagab lõpliku prognoosi; (ii) elamute energia nõudluse prognoosimudelist (LSTM/XGBoost); (iii) EHS optimeerijast, mis on realiseeritud heuristiliste reeglitena (deterministlik MILP Pyomo/GLPK mudel) ja stsenaariumipõhise stohhastilise MILP-na, et ilmutatud kujul käsitleda prognoosi ja tariifi määramatust.

100 hindamispäeva jooksul ületas hübriidprognoos LSTM, BiLSTM, XGBoost, LightGBM, encoder–decoderi, füüsilise mudeli ja püsiva baasjoone tulemused, kinnitades hüpoteesi, et hübriidmudel parandab tulemuse täpsust puudulike andmete ja ilmastikust sõltuva konfiguratsiooni korral.

Tehno-majanduslik raamistik ühendab turuhinnad, inflatsiooni, päikesepaneelide tootlikkuse mudeli ja aku vananemismudeli, et arvutada realistlikud tasuvusajad. Reaalseid elamute andmeid kasutavad uuringud näitavad, et PV pargi võimsus määrab kasumlikkuse, samas kui aku mahutavus mängib teisejärgulist rolli; PV-süsteemid > 9 kWp annavad tavaliselt lühima tasuvusaja, samas kui alamõõdulised süsteemid võivad olla majanduslikult ebaotstarbekad.

Riistvaraga seoses pakub väitekiri välja SC-TP EnergyR topoloogia, mis suudab rakendada ühte muunduri elementi (mitte samaaegselt) kolme faasi jaoks kasutades relee juhtimist, võimaldades nii faaside vahelist võimsuse jaotamist ja tasakaalustamist kodus ja võrgu liidesel. See topoloogia vähendab faaside tasakaalustamatust, suunates PV/aku võimsuse kõige koormatunasse faasi, vältides samal ajal kaht lisamuunduri elementi.

Kogu platvorm on eksperimentaalselt valideeritud TalTechi liginullenergiahoones. Kokkuvõttes töötati välja külmkäivitust taluv päikeseenergia prognoosimise meetod, stsenaariumipõhine stohhastiline EHS optimeerija, akude võimsuse vähenemisega arvestav infrastruktuuri mõõtmise tööriist ja SC-TP EnergyR kontseptsioon, mis vähendab kapitalikuludid ja parandab tehnilist jõudlust. Need saavutused edendavad elamute valmisolekut osaleda tulevikus intelligentsetes, taastuvenergia poolt domineerivates elektrisüsteemides.

Appendix

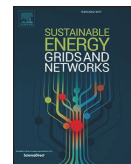
Publication I

H. N. Hokmabad, O. Husev, J. Kurnitski, and J. Belikov, 'Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications', *Sustainable Energy, Grids and Networks*, vol. 38, p. 101385, 2024.



Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

journal homepage: www.elsevier.com/locate/segan

Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications

Hossein Nourollahi Hokmabad^{a,b,*}, Oleksandr Husev^a, Jarek Kurnitski^c, Juri Belikov^b

^a Department of Electrical Power Engineering & Mechatronics, Tallinn University of Technology, Tallinn 19086, Estonia

^b Department of Software Science, Tallinn University of Technology, Tallinn 11712, Estonia

^c Department of Civil Engineering and Architecture, Tallinn University of Technology, Tallinn 19086, Estonia

ARTICLE INFO

Keywords:

Battery energy storage
Battery degradation
Optimal sizing
Energy management
Techno-economic analysis
Rule-based non-linear optimization

ABSTRACT

Battery energy storage systems (BESSs) are essential in enhancing self-sufficiency, sustainability, and delivering flexibility services. However, adoption of this technology in residential applications is constrained, predominantly due to its suboptimal economic performance. A proper selection of design parameters and optimal resource utilization can significantly enhance performance and establish economic feasibility. This research introduces a photovoltaic (PV)-BESS optimization framework, formulated to ascertain optimal infrastructure sizing, and maximize economic performance. The proposed tool considers various elements, such as energy storage state of health status, renewable profiles, residential load profiles, and prevailing energy market conditions. A novel rule-based, non-linear optimization method is developed, with a focus on maximizing revenue while considering energy storage (ES) degradation to project more accurate and realistic scenarios and payback periods. Employing the Estonian energy market as an illustrative case study, the model probes the potential of integrating ES and PV technology in North European residences. The model achieved the shortest payback period of five years in our case study, underscoring the emerging potential of such technologies in behind-the-meter applications.

1. Introduction

THE profound impacts of carbon emissions, including global warming and climate change, are currently both undeniable and intolerable. Fossil fuels have been the predominant source of such emissions for centuries. This has led to the intensified adoption of stricter global regulations and the development of eco-friendly technologies aimed at mitigating reliance on fossil fuels. For example, the European Union's energy performance of buildings directive necessitates that all new constructions adhere to near-zero energy building (nZEB) standards [1]. This regulation implies that newly constructed buildings are obligated to produce as much energy as they consume on an annual scale.

Solar panels, or PVs, facilitate the harvesting of solar energy, with capabilities ranging from several watts to megawatts. This technology

stands out in the residential sector compared to other green energy methods, owing to its adaptable scale, low-maintenance nature, and extended lifespan. However, the intermittent nature of solar energy requires innovative solutions to fully harness its benefits. For example, integrating energy storage systems (ESSs) into existing PV setups serves as a significant solution. Such storage systems alleviate the intermittency of renewable energy sources, stabilize grid frequency, and contribute to enhanced flexibility and power balancing [2]. Furthermore, the substantial gaps in electricity tariffs have made energy arbitrage a profitable earning opportunity for ESSs [3].

Despite the numerous advantages of including energy storage systems beside PV setups, their adoption has not piqued public interest, largely due to economic drawbacks, such as high upfront costs and long payback periods [4], [5]. In many regions without subsidies, the economic viability of integrating ESSs is often questioned [6].

Abbreviations: BESS, Battery energy storage system; PV, Photovoltaic; ES, Energy storage; SoH, Energy storage state of health; nZEB, Near-zero energy building; ESS, Energy storage system; LP, Linear programming; MILP, Mixed integer linear programming; FiT, Feed in tariff; CC, Capital cost; UT, Utility tariff; ToU, Time-of-use; ER, Energy Router; EMU, Energy management unit; SoC, State of charge; DoD, Depth of discharge; FT, Flat tariff; SP, Seasonal pricing; DT, Demand tariff; EoL, End-of-life; EV, Electrical vehicle.

* Corresponding author at: Department of Electrical Power Engineering & Mechatronics, Tallinn University of Technology, Tallinn 19086, Estonia

E-mail address: hossein.nourollahi@taltech.ee (H.N. Hokmabad).

<https://doi.org/10.1016/j.segan.2024.101385>

Received 14 October 2023; Received in revised form 29 February 2024; Accepted 9 April 2024

Available online 16 April 2024

2352-4677/© 2024 Elsevier Ltd. All rights reserved.

Consequently, various governments have implemented subsidy incentives to propel the adoption of these technologies [7], [8]. For instance, the newly implemented tax reduction policy for energy storage installations in Italy significantly impacted the economic viability and profitability of the technology [6]. On the other hand, integration of PV systems into residential grids is currently proven to be highly profitable worldwide [9]. Thus, incentives which only support PV installations are scheduled to be phased out soon.

Projection of utility prices for the next 20 years indicates an upward trend due to increased demand, transition to renewable energy sources, and infrastructure investments [4]. Fig. 1 illustrates the weekly average end-user electricity price trends in Estonia from 2020 to 2022. Statistics reveal that, at several points in 2022, energy costs were up to ten times higher than in previous years, setting new records. Additionally, the utility tariff experiences notable daily fluctuations. Fig. 2 displays the oscillations in electricity tariffs on August 23rd, 2022, with the peak period price being approximately five times that of the off-peak period on the same day. These daily fluctuations in tariffs introduce new opportunities for energy arbitrage and underscore the importance of load shifting and peak shaving in reducing consumers' electricity bills. Such growing and disproportionate trends signal to users the need to optimize their consumption patterns, both quantitatively and temporally.

Existing studies will soon be less accurate due to new market variables, policy adjustments, and the advent of new technologies. However, it remains crucial to incorporate the findings and methodologies proposed in these studies for future research. Paper [10] examined the life cycle cost optimization of BESS in residential applications, concluding that for a standalone system, the break-even price for Li-ion batteries is 400 €/kWh. In contrast, for grid-connected modes, the calculated BESS price is as low as 30 €/kWh, which is significantly below the current market prices.

The authors in [11] calculated the optimal size of the hybrid PV-BESS system using commercially available products like TESLA and ENPHASE power walls, concluding that BESS prices would need to decrease by 40–70% to be profitable for customers. Work [12] analyzed various electricity pricing methods and, in the best scenario, determined the minimum payback period to be approximately 10 years. Researchers in [13] explored the integration of BESS in existing PV systems, evaluating three different tariff strategies including fixed, dynamic, and double tariff. Another study [14] found that utilizing BESS with grid-connected PV systems is economically viable for residential buildings in Finland. Validation of simple rule-based methods for BESS control and battery degradation model have been investigated with experimental measurements in [5]. Study [4] identified that, for average homes in Germany, the optimal configuration comprises a PV system size above 6 kWp combined with a 5 kWh BESS and a 2.5 kW inverter.

Control strategies and optimization methods play an essential role in concluding the economic model. In [15], ESSs control methods have been reviewed. The Internet of energy-based building energy management systems is reviewed in [16]. Optimal charge/discharge

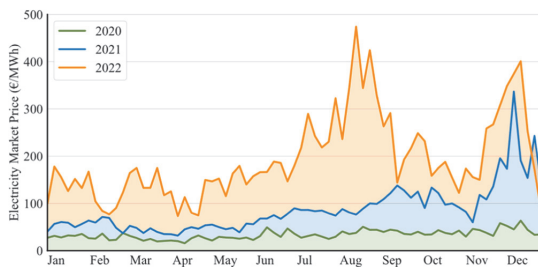


Fig. 1. Weekly average electricity tariff in Estonian energy market.

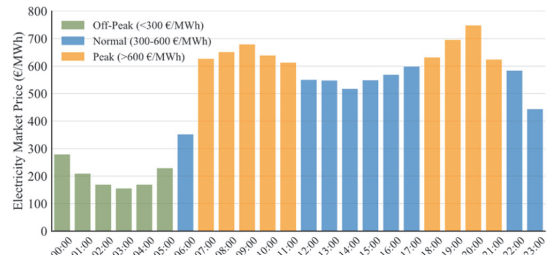


Fig. 2. Electricity price oscillations on August 23rd, 2022, Estonia.

scheduling for BESS considering feed-in feasibility has been introduced in [17]. Work [18] proposed a MILP optimization method for a home energy management system with the main objectives of reducing utility cost, demand side management, and maximizing occupant comfort levels. HOMER, a microgrid software design tool, has been used to obtain optimum sizes for hybrid PV-BESSs in [19,20]. The multi-objective optimization framework has been developed in [21]. The authors considered energy and power independence and payback period as objectives and feed in tariff (FIT), battery degradation, PV orientation, and setup capital cost (CC) as optimizing variables. The operation and optimization of grouped BESS have been evaluated in [22]. The interaction of electric vehicles (EVs) as an ES unit for PV-integrated houses has been evaluated in [23,24]. In [2] a residential apartment is analyzed as a local energy community for a cost-effective PV-BESS design. Conventional and intelligent energy management methods have been reviewed in [25].

Given that Li-ion batteries possess a typical operational lifespan of less than 15 years, the majority of cited studies have concluded that, without governmental assistance and subsidies, the installation of residential BESS remains economically unviable. However, considering recent significant shifts in the energy market such as reductions in battery costs, increased electricity prices, and technological advancements, the integration of energy storage with PV systems seems to be gaining interest due to the potential for profitable returns. Nonetheless, economic viability is profoundly influenced by the system's design process and operational policy.

Many studies have incompletely addressed various technical aspects in their models. Some have omitted considerations of aging factors [26], [27], [28] including battery [6], [19], [29] and solar panel degradation [29], [30]. Others have overlooked the impacts of power electronics, accounting for storage round trip efficiency [14], [22], maintenance [12], [13], [22], and replacement costs [31]. Additionally, economic elements like annual inflation were ignored in [13], [21], [22]. Some studies employed optimistic prices [32], and many incorrectly treated elements' size/cost coefficients as constant. Consequently, the results from these studies may tend to be optimistic. Including such non-ideal parameters is, however, unnecessary when even the ideally modeled system lacks economic attractiveness. Table 1 compares the proposed solution with the state-of-the-art and provides an overview of the objectives and methods applied in the literature.

The lack of reliable personalized techno-economic models has left end-users in uncertainty regarding the profitability of deploying such systems. A comprehensive personalized model that encapsulates all the mentioned details can thus offer clearer insights for customers and authorities. Additionally, a universal platform applicable to diverse situations across different regions could facilitate the discovery of optimal policies and enable the designing of personalized systems suited to individual cases.

Addressing identified gaps, this paper presents a personalized techno-economic model evaluating diverse parameters on system profitability and payback time. The model integrates a physics-based battery SoH estimator, a linear PV degradation model, and an online tariff

Table 1
Literature review and comparison with current study.

Ref.	Location	Year	Aging factors	Energy autonomy	Energy self-consumption	Optimization goal	Optimization method	Uncertainty	Online energy market data
PP	Estonia	2022	✓	✓	✓	Profit maximization and storage fading minimization	Rule-based and none-linear	×	✓
r [26]	China	2020	×	×	Profit maximization	Rule-based	✓	✓	✓
r [33]	Germany	2014	×	×	Cost minimization	Linear programming (LP)	×	×	✓
r [34]	Australia	2019	✓	✓	Cost minimization	Model predictive control	✓	✓	✓
r [35]	UK	2022	✓	×	Payback minimization	LP	×	×	✓
r [36]	Turkey	2022	✓	×	Profit maximization	Mixed Integer LP (MILP)	×	×	×
r [37]	India	2021	✓	×	Profit and comfort maximization	MILP	×	×	×
r [27]	USA	2020	×	×	Reliability while outage	MILP	✓	✓	✓
r [18]	India	2020	✓	×	Bill minimization	MILP	✓	✓	✓
r [38]	Australia	2021	✓	×	Net present cost reduction	Rule-based	×	×	×
r [39]	Tunisia	2018	×	×	Power flow reliability	Artificial Intelligence	×	×	×
r [40]	Iran	2020	×	×	Bill minimization	MILP	✓	×	×
r [41]	China	2022	×	×	Net present cost reduction	Rule-based	✓	×	×
r [32]	Estonia	2022	×	✓	Self-consumption maximization	Analytical scheduling	×	✓	✓
r [42]	Thailand	2022	✓	×	Net present cost reduction	Homer Pro.	×	×	×
r [43]	Thailand	2022	✓	✓	Reliability maximization	Rule-based	×	×	×
r [44]	China	2022	✓	✓	Life cycle cost minimization	MILP	×	×	✓
r [45]	UK	2021	×	×	Payback time minimization	Rule-based	×	×	✓
r [46]	China	2021	×	×	Total cost minimization	MI nonlinear programming	×	×	✓
r [47]	Australia	2019	×	×	Cost minimization	Adaptive robust optim.	✓	×	×

extractor. It also factors in key economic variables, such as energy market dynamics, inflation rates, and projections. By using a novel rule-based non-linear objective function, the model optimizes energy distribution in the residential grid to both maximize annual revenue and minimize storage degradation. This involves solving a cost function in every iteration to achieve the optimal energy distribution and shortest payback period. While this research focuses on the Estonian market, the adaptable model can cater to different regions and residential setups with parameter adjustments and requisite data input.

2. Energy market and system structure

Estonia, along with Norway, Sweden, Finland, Denmark, Latvia, and Lithuania, belongs to Nord Pool’s open electricity market, where electricity prices are predetermined by retailers 24 hours in advance. Eesti Energia, the primary energy supplier in Estonia, offers four tariff packages to its residential customers, summarized in Table 2. In this study, it is presumed that the customer has chosen the “Exchange” package for their property, and the utility tariff (UT) is calculated based on time-of-use (ToU) prices. Additionally, the Estonian government has introduced a new incentive for residential solar energy providers. Consequently, starting from the second half of 2022, the FiT is set to be 20% below the ToU electricity purchase price.

Load demand and PV generation profiles were collected in 2022 from a residential property in Tallinn, Estonia. Due to the capacities of the power electronic devices, the peak load demand power in this study is established at 5 kW. Fig. 3 illustrates the solar production, categorized into four different days across different seasons with a 5-minute resolution. Given Estonia’s geographical location, the country experiences extended daylight in the spring and summer and shorter days in the autumn and winter.

The abstract structure of PV-BESS equipped residential grid is shown in Fig. 4. A universal dc/dc/ac inverter introduced in r [48], with the peak efficiency of 98%, has been selected as the conversion unit between all parties. In this paper, we refer to this device as the energy router (ER), which is pivotal in managing smart meter tasks within our platform. The ER is tasked with monitoring internal grid statuses, including line voltage, current, and phase differences. It subsequently relays this grid status data to the upper-level management system. The energy management unit (EMU) is responsible for gathering energy market data from web servers or databases, as well as the statuses of energy storage systems. Utilizing both the provided and collected data, the EMU’s algorithm generates control signals. These signals are then dispatched back to the ER for implementation within the residential grid, ensuring an optimized and efficient energy distribution.

2.1. BESS specifications

The selected lithium-ion (Li-ion) cells for this application are of the lithium iron phosphate (LiFePO4) type, part of the 18650 Li-ion battery family. This technology is recognized for its suitability in residential and EV applications due to its safety, long cycle life, and thermal stability, as highlighted in reference r [49]. To maximize the lifespan of these battery cells, operational guidelines based on the manufacturer’s recommendations dictate that the maximum and minimum allowable state of charge (SoC) levels should be maintained at 90 % and 15%, respectively. Adhering to these SoC thresholds ensures that the maximum possible

Table 2
Utility tariffs in estonia for residential clients.

Package title	Price methodology	Price	Unit
Universal service	Changes every month	19.24	cents/kWh
Fix plus	Fixed green kWh price	21.00	cents/kWh
Fix	Fixed	31.70	cents/kWh
Exchange	Changes every hour	28.22	cents/kWh

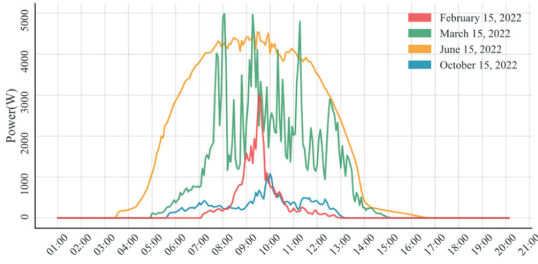


Fig. 3. Solar power generation at different days selected from four different months in 2022, Tallinn, Estonia.

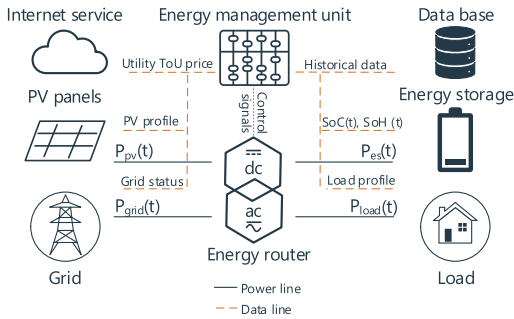


Fig. 4. Energy management unit abstract structure equipped with PV-BESS-ER.

incremental depth of discharge (DoD) — the measure of how much energy has been drained relative to the battery’s total capacity — does not exceed 75 %. The difference between the initial and final SoC during a cell’s discharge cycle, denoted as ΔDoD , is critical for optimizing battery health and performance over time. This operational strategy not only prolongs the life of the battery but also maximizes energy availability for user applications.

Normally 20 % reduction from battery nominal capacity is considered as the cells’ end-of-life (EoL) point for EVs and mostly this criterion is selected between 20 % and 40 % for residential applications [50]. In this study, battery cells EoL is considered as a point where cell capacity reaches 70% of its nominal capacity. This criterion enables the re-use of second-hand EV battery packs in residential grids. Some studies [51], [52], [53] pointed the possibility of benefiting secondhand retired EV batteries as energy storages. Subsequently, researchers have investigated the application of second-hand batteries in residential grids [35], [54]. However, batteries during their second life have lower roundtrip efficiency due to their fast internal resistance growth. This phenomenon decreases the system’s overall performance and leads to higher power losses.

Table 3
Bess Technical Specifications.

Specification	Value	Unit
Cell voltage	3.3	V
Cell stored energy	3.63	Wh
Cell nominal capacity	1.1	Ah
Internal resistance	12.6	mΩ
Battery roundtrip efficiency	92	%
Continuous discharge current	30	A
Continuous charge current	4	A
ΔDoD	75	%
EoL	70	%
Cycle life, 100% DOD	2000>	Cycles
Calendar life	10	Years

The specifications for the cells of the battery pack can be found in Table 3, with most of the parameters extracted from the battery datasheet [55]. Energy storage is presumed to be housed within dwellings. Consequently, the ambient temperature is maintained at 25 °C year-round, irrespective of the external temperature. For the sake of simplicity, it is assumed that the temperature is uniform across all battery cells. Based on the datasheet values and cell current ratios, the temperatures of all cells are presumed to be 30 °C during operation. This assumption is valid as the single-cell current during the controlled and limited charge/discharge phases is significantly below the maximum permissible current limits. Note that, to meet the required energy storage capacity in the designed residential grid, battery cells are stacked in series and parallel form. For instance, in order to have a BESS with 5 kWh energy storage capacity and selecting an output dc voltage level equals to 330 V, 100 battery cells should be stacked in series and in 14 parallel branches.

3. Proposed method

In this section, the proposed method will be introduced, starting with the BESS degradation model. This will be followed by the development of the system cost function and the formulation of the optimization problem. Then, the energy management algorithm will be discussed. Finally, the economic model and the principles of calculation will be discussed in detail.

3.1. Battery degradation model

Many studies have largely overlooked the impacts of battery degradation on model performance. This simplification may be a valid approach for systems that are inherently unprofitable. However, for models that are potentially cost-effective, it is presumed essential to consider the aging impacts of Li-ion batteries. Different methods of battery usage significantly influence the rate of battery aging. Consequently, optimal strategies can effectively prolong battery life and favorably influence the economic model’s performance. The effects of BESS degradation in various systems have been explored in [56], [57], [58]. Several works have assessed the degradation characteristics of Li-ion batteries and their influencing parameters. Data-driven approaches have been employed in [59], [60], [61]. Physics-based models are discussed in [62], [63], and reduced physics-based solutions for SoH estimation have been utilized in [64], [65]. Finally, equivalent circuit models have been implemented in [66], [67].

Capacity fading is the most significant drawback of Li-ion battery cell technology. This parameter refers to remained cell capacity in comparison with nominal cell capacity. This phenomenon is specified with two major factors: degradation due to time pass and fading due to the amount of processed charge [49]. As a result, total capacity fading $C_{fd} \in [0, 100]\%$ can be obtained as:

$$C_{fd} = C_{cal,fd} + \frac{C_{cycle,fd}}{C_{nm}} \times 100, \quad (1)$$

where $C_{cal} \in [0, 100]\%$ represents the percentage of calendar capacity loss from the initial nominal capacity due to battery stock or long-term relaxation and $C_{cycle} \in [0, C_{nm}]$ is the capacity degradation resulting from the amount of the processed charge during cells recharge/discharge cycles and $C_{nm} \in (0, 15]$ kWh is the rated energy of the fresh battery cell. Finally, by considering both degradation factors, the $SoH \in [0, 100]\%$ level of the energy storage can be formulated as:

$$SoH = 100 - C_{fd}. \quad (2)$$

For fresh batteries C_{fd} is zero and SoH is equal to 100%. Thus, SoH equals to zero indicates that all capacity of the energy storage is decayed, and the battery is no longer capable of storing and delivering energy.

In the developed energy storage fading model, calendar capacity loss is applied using the superposition rule and is calculated as τ [68]:

$$C_{cal,m} = (0.019 \times SoC^{0.823} + 0.5195) \times (3.258 \times 10^{-9} \times T^{5.087} + 0.295) \times m^{0.8}, \quad (3)$$

where $C_{cal,m} \in [0, 100]\%$ represents the percentage of total calendric capacity loss over the course of m months, and T denotes the ambient temperature in degrees centigrade, and SoC refers to the cells' available charge level, expressed in percentage, during batteries' stock period.

In order to calculate the calendar loss during energy storage operation, Eq. (3) should be reformed. For simplification, the SoC value is approximated to be constant at 1-hour intervals. As a result, the SoC value is averaged for the corresponding time slot, represented by $SoC_{1h,avg} \in [0, 100]\%$. Given that the cell temperature is maintained at 25 °C during relaxation times, the capacity fading ratio for 1 hour period represented by $C_{cal,1h} \in [0, 100]\%$ can be computed as:

$$C_{cal,1h} = (6.6148 \times SoC_{1h,avg} + 4.6404) \times 10^{-6}. \quad (4)$$

In Eq. (4), $C_{cal,1h}$ represents the percentage of the capacity loss during each hour of storing battery cells in the mentioned condition and $SoC_{1h,avg}$ is the average of SoC levels during each hour. So, it can be concluded that a higher SoC ratio accelerates calendar ageing and for a long-term stock, cells should not be fully charged. Finally, the $C_{cal,fd}$ can be linearly expressed as:

$$C_{cal,fd} = \sum_{\tau=1}^h C_{cal,1h}[\tau], \quad (5)$$

where h is the total hours of energy storage operation and relaxing times.

To calculate the capacity fading of cells due to charge circulations, the model proposed in τ [65] has been applied. In this empirical model, energy storage degradation is calculated based on SoC level average and its deviation from the average value during each charge/discharge event. The total cyclic degradation is compiled as:

$$C_{cycle,fd} = \sum_{i=1}^e F_i \times Q_i \times e^{-\frac{E_a}{R T_i}} \left(\frac{1}{T_i} - \frac{1}{T_{ref}} \right), \quad (6)$$

$$F_i = k_{s1} SoC_{dev,i} e^{k_{s2} SoC_{avg,i}} + k_{s3} e^{k_{s4} SoC_{dev,i}}, \quad (7)$$

where $C_{cycle,fd}$ is calculated by accumulating all recharge and discharge events denoted as i . Q_i is the amount of processed charge during each recharge/discharge event which is function of cell's current and time duration of passed current. Here, $k_{s1}, k_{s2}, k_{s3}, k_{s4}$ are the parameters of the empirical capacity fading model. E_a is the activation energy, R is the gas constant, T_i is the cell temperature at the i^{th} charge or discharge moment, T_{ref} is a reference temperature, both in Kelvin, and e is the total number of recharge and discharge events. The numerical values for the mentioned parameters are collated in Table 4. In Eq. (7), SoC_{avg} is average amount of SoCs and SoC_{dev} is normalized standard deviation of

SoC from SoC_{avg} during each charge/discharge operation. For instance, if a full-charged cell is completely discharged in one cycle, then both SoC_{avg} and SoC_{dev} will be obtained as 50%. SoC_{avg} and SoC_{dev} have been formulated as:

$$SoC_{avg} = \frac{1}{\Delta Q_i} \int_{Q_{i-1}}^{Q_i} SoC(Q) dQ, \quad (8)$$

$$SoC_{dev} = \sqrt{\left| \frac{3}{\Delta Q_i} \int_{Q_{i-1}}^{Q_i} (SoC(Q) - SoC_{avg})^2 dQ \right|}, \quad (9)$$

where Q_i is the final amount of processed charge, Q_{i-1} is the initial amount of processed charge before starting of energy circulation through battery cells, and ΔQ_i is the absolute amount of charge processed in each operation event. It should be mentioned that this model holds accuracy only under conditions where the cells are persistently operating at temperatures above 25 °C.

3.2. System operation optimization problem

For optimized energy circulation, there is a need for defining a proper cost function based on optimization goals to select the optimum strategy for utilizing resources and demand responding. This optimization problem should be continuously solved during system operation based on new conditions and subjections. Here, maximizing profit considering energy storage capacity fading, has been defined as the optimization problem's main objective. Thus, the system cost function is formulated as:

$$\min_x \mathcal{F}^T \bullet \mathcal{Z} \rightarrow \min \sum_{j=1}^n Gr_{c,j} + Es_{c,j} - F_{p,j}, \quad (10)$$

where $Gr_c \in \mathbb{R}^n$ is the utility usage cost, $Es_c \in \mathbb{R}^n$ is the cost of energy storage fading, $F_p \in \mathbb{R}^n$ is the amount of profit system can earn by trading energy with the utility grid, n is the total optimization iterations steps, and j represents j^{th} iteration. In Eq. (10), $\mathcal{F} \in \mathbb{R}^l$ is the vector of optimization variables, and $\mathcal{Z} \in \mathbb{R}^l$ represents the total optimization factors,

$\vec{\mathcal{Z}} = [\vec{P}_{PV/G}, \vec{P}_{PV/L}, \vec{P}_{PV/ES}, \vec{P}_{G/L}, \vec{P}_{G/ES}, \vec{P}_{ES/L}, \vec{P}_{ES/G}]^T$ where, suffix G represents utility grid, ES represents energy storage, PV represents photovoltaic setup, and L represents demand or load. $P_{x/y} \in \mathbb{R}^n$ represents the amount of power flow from point x to point y , and l represents the total dimension of the optimization vector. In Estonian energy market, $Gr_{c,j}$ for residential sector is calculated as:

$$Gr_{c,j} = E_{G,j} \times U_{ToU,j} + C_{fx} + E_{Ex,j} \times C_{Ex}, \quad (11)$$

where $\vec{E}_G \in \mathbb{R}^n$ is the net amount of electricity purchased from the main grid in each iteration, $U_{ToU} \in \mathbb{R}^n$ is vector of electricity tariff in each corresponding iteration, $C_{fx} \in \mathbb{R}$ is the network service charge which is currently 2.02 €/month, $E_{Ex} \in \mathbb{R}^n$ is the total amount of energy exchange with the grid in each iteration. E_{Ex} is calculated accumulating total amount of imported and exported energy, and $C_{Ex} \in \mathbb{R}$ is the fixed cost of energy exchange with the grid and is 0.45 cents/kWh.

Energy storage fading expenses Es_c , is calculated based on the capital cost of BESS installation and assuming that when the EoL point is met, the storage will not have any value for the user. Thus, the $Es_{c,j}$ is calculated as:

$$Es_{c,j} = -U_{BESS} C_{nm} \left(\frac{\Delta SoH_j}{100 - EoL} \right), \quad (12)$$

where U_{BESS} €/kWh is the unit price of energy storage installation. Thus, $U_{BESS} C_{nm}$ is the total cost of adding energy storage to the dwelling. $\Delta SoH \in \mathbb{R}^n$ is the amount of degradation that occurs due to energy

Table 4
Circular capacity fading model's constant parameters.

Parameter	Value	Unit
k_{s1}	-4.092e-4	-
k_{s2}	-2.167	-
k_{s3}	1.408e-5	-
k_{s4}	6.13	-
E_a	78.06	k.mol/J
R	8.314	J/k.mol
T_{ref}	298.15	K
T_i	303.15	K

storage utilization in each optimization stage and ΔSoH_j is equal to $SoH_j - SoH_{j-1}$. Consequently, the cost associated with utilizing energy storage rises as SoH decreases. Finally, $F_{P,j}$ is computed by:

$$F_{P,j} = (E_{f,j} \bullet U_{ToU,j}) \times \gamma, \quad (13)$$

where $E_f \in \mathbb{R}^n$ is the array of total amount of dropped energy to the grid in each optimization iteration, and $\gamma \in [0, 1]$ is a FiT coefficient.

Indeed, energy is the time integral of power, and under the assumption of constant time intervals and constant system voltage and current levels within each interval, energy parameters can be represented by power multiplied by time values $E = P \times t$. So, the optimization function can be reformulated and solved based on residential distribution grid's power values. For each infrastructure configuration the constraints will be different based on PV, BESS, and ER size. For instance, considering PV, BESS, and ER as 3 kWp, 5 kWh, and 5 kW respectively, the non-equality constraints for this particular setup will be as:

$$\forall j, 0 \leq P_{PV/G}[j], P_{PV/L} [j], P_{PV/ES}[j] \leq 3000, \quad (14)$$

$$\forall j, 0 \leq P_{ES/G} [j], P_{ES/L}[j], P_{G/ES}[j], P_{G/L}[j] \leq 5000, \quad (15)$$

$$\forall j, P_{G/L}[j] + P_{ES/L}[j] + P_{PV/L}[j] \leq 5000, \quad (16)$$

$$\forall j, P_{ES/G}[j] + P_{PV/G}[j] \leq 5000, \quad (17)$$

$$\forall j, P_{ES/G}[j] + P_{ES/L}[j] \leq 5000, \quad (18)$$

$$\forall j, (P_{ES/L}[j] + P_{ES/G}[j]) \times t \leq (SoC[j] - SoC_{\min}) \times C_{nm} \times SoH[j], \quad (19)$$

$$\forall j, (P_{PV/ES}[j] + P_{G/ES}[j]) \times t \leq (SoC_{\max} - SoC[j]) \times C_{nm} \times SoH[j], \quad (20)$$

$$\forall j, P_{PV/ES}[j] + P_{G/ES}[j] \leq 5000, \quad (21)$$

$$\forall j, P_{PV/G}[j] + P_{PV/L} [j] + P_{PV/ES}[j] \leq 3000, \quad (22)$$

$$\forall j, P_{G/ES}[j] + P_{G/L}[j] \leq 5000, \quad (23)$$

where, suffix G represents utility grid, ES represents energy storage, and L represents demand or load. $P_{x/y}[j]$ represents the amount of power flow from point x to point y in the house distribution system during j^{th} iteration. $C_{nm} \times SoH[j]$ represents the amount of remaining capacity inside the BESS. In Eq. (19) and Eq. (20), t represents the iteration intervals or steps in hours. For instance, the proposed optimization engine is designed to optimize the power flow every 5 minutes.

The equality constraints are listed as follows:

$$\forall j, P_{PV/G}[j] + P_{PV/L} [j] + P_{PV/ES}[j] = P_{PV,\text{total}}[j], \quad (24)$$

$$\forall j, P_{PV/L} [j] + P_{ES/L}[j] + P_{G/L}[j] = P_{L,\text{total}}[j], \quad (25)$$

where $P_{PV,\text{total}}$ is the total amount of power generated by solar cells and $P_{L,\text{total}}$ is the total amount of demand or load in the corresponding time interval. Nonlinear equations are introduced as:

$$\vec{P}_{G/ES} \bullet \vec{P}_{ES/G} = 0, \quad (26)$$

$$\vec{P}_{PV/ES} \bullet \vec{P}_{ES/G} = 0, \quad (27)$$

$$\vec{P}_{PV/ES} \bullet \vec{P}_{ES/L} = 0, \quad (28)$$

$$\vec{P}_{G/ES} \bullet \vec{P}_{ES/L} = 0. \quad (29)$$

These limitations prevent the events in which the algorithm may decide to charge and discharge the energy storage simultaneously. Finally, $E_{f,j}$ and $E_{G,j}$, can be compiled as:

$$E_{f,j} = (P_{ES/G}[j] + P_{PV/G}[j]) \times t, \quad (30)$$

$$E_{G,j} = (P_{G/ES}[j] + P_{G/L}[j]) \times t. \quad (31)$$

The pre-defined nonlinear optimization problem was addressed using the MATLAB optimization toolbox and the 'fmincon' solver. This solver utilizes a gradient of the cost function to find the solution corresponding to objective function global minimum.

After solving the initial power flow optimization problem which defines best energy flow strategy for each particular configuration, we employed a grid search method for selecting the optimal setup configurations for each particular PV size, as the size variables are integers and the solver cannot handle these in its optimization algorithm, thus another stage of optimization is required.

In determining the search space for PV, BESS, and ER sizing, we surveyed available market products. Batteries up to 15 kWh are commonplace for residential energy management. The primary determinant for PV systems is installation area; thus, a 20 kWp system, occupying approximately 80 m², was chosen as the maximum size. ER size varies based on PV and storage capacities, with numerous options available. Table 5 summarizes the chosen variable ranges. Furthermore, all possible PV-BESS-ER combinations are simulated under two different conditions: availability of energy trade with the grid and absence of possibility to export energy to the grid, to identify the optimal energy storage and ER sizes for each specified PV size in the two possible scenarios. Moreover, during system operation, the C-rate ratios for the BESS charging and discharging procedures are confined to the manufacturer's recommended values, which are detailed in Table 5. Additionally, the ER maximum power ratio plays a crucial role in determining the maximum power that can be delivered to or received from the BESS. For example, if the ER size is set at 3 kW, the maximum achievable charge or discharge ratio for a 5 kWh BESS would be 0.6.

Including forecasting data can enhance the performance of EMUs. However, the designed cost function and the core optimization engine do not incorporate forecasts into their optimization principles. This limitation leads to decisions being made based on the current system state, restricting the benefits from future knowledge. To address this issue partially, a rule-based decision-making block has been added to the main optimization loop. The rule-based block utilizes day-ahead energy and load profile forecasts to generate charge and discharge signals for energy storage when certain conditions are met. Thus, at the beginning of each day, total solar energy generation and total load demand for the next 24 hours are calculated. Additionally, the minimum and maximum electricity prices and corresponding hours are extracted from the energy market.

In moments with minimum energy prices, if one of the conditions below is met, the block sends a charge command to the energy storage:

1. The forecasts for total generated energy should be less than the forecasts of total load demand (This confirms that the PV will not be able to handle all energy demand solely). In this stage, the battery can be charged in the amount to fill the existing gap.

Table 5
Simulated variables ranges and steps.

Variable name	Range	Step	Unit	Symbol
BESS size	1–15	1	kWh	\overline{BESS}_{size}
PV size	1–20	1	kWp	\overline{PV}_{size}
ER size	2.5–20	2.5	kW	\overline{ER}_{size}
C-rate charge	0–1.35 C	Continuous	-	$C_{rate, ch}$
C-rate discharge	0–1 C	Continuous	-	$C_{rate, dis}$

2. $PR_{\max} - PR_{\min} \geq Es_{c,j}$, where PR_{\min} and PR_{\max} are the minimum and maximum prices of electricity on the following operation day, respectively (This ensures that the cost of utilizing energy storage is less than the benefit from energy arbitrage). In this condition, the energy storage will be fully charged to maximize energy arbitrage.

In peak periods, if both of the conditions below are met, the block sends a discharge command to the energy storage:

1. The stored energy in the battery plus the energy generation forecast should be higher than the load demand forecasts for the rest of the day (This ensures the energy storage has surplus energy for trading). In this condition, the energy storage is allowed to release surplus energy to the grid.

2. $PR_{\max} - PR_{\min} \geq Es_{c,j}$.

In this study, ideal forecasts are utilized for simplification; however, incorporating real forecast data will tie the effectiveness of the decisions to the accuracy of the forecasts.

3.3. Energy management framework

Fig. 5 illustrates the overview of the proposed energy management platform. Initially, the platform collects all available setup configurations, real-time grid tariffs, economic parameters, house consumption profiles, solar energy generation profiles, system technical parameters, and energy storage status. For each candidate configuration, the system runs simulations for 20 years to model the energy circulation inside the residential grid based on the inputs and system specifications. In every iteration, the optimization engine determines the optimum method to leverage available resources to meet the demand and attain the defined goals. If feasible, after proper load management, the device exchanges energy with the main grid to accrue benefits.

During each iteration, the platform calculates the energy processed through energy storage to update SoC and SoH parameters for the subsequent iteration. Once all possible combinations have been computed, the economic model calculates system revenue and payback time. Based on these calculated values, the optimal configuration for each PV size is selected and reported. Additionally, user-related reports such as total energy production, battery health status, system performance, and others are documented at this stage. It's critical to acknowledge that

when the energy storage SoH level declines below the EoL point, the model anticipates that the energy storage will be replaced with new batteries. Replacement also occurs for ER after 15 years of operation, substituting the old device with a new one. Nevertheless, it is assumed that the solar panels will work properly during the operation period. Therefore, there is no need to replace the PV panels, yet 0.5% annual linear performance degradation denoted as Ω_{pv} is implemented in the calculations [33].

PV self-consumption, and energy autonomy factor are two fundamental factors for evaluating the nZEB and EMU performance. However, maximizing these factors has not been considered as an optimization goal in the proposed method, yet the system's performance for both factors has been evaluated and reported. In this paper, self-consumption denoted as $\delta \in [0, 100]\%$, is defined as follows:

$$\delta = \frac{E_{pv/l}}{E_{pv}} \tag{32}$$

where, $E_{pv/l} = \sum_{j=1}^n E_{pv/l,j}$ is the total amount of delivered PV energy to the load, and $E_{pv} = \sum_{j=1}^n E_{pv,j}$ is the total amount of green energy production. Furthermore, the energy autonomy factor denoted as $\epsilon \in [-1, 100]\%$, indicates the percentage of the imported energy from the main utility grid in comparison with the amount of load demand, is calculated as [21]:

$$\epsilon = \frac{E_l - E_{grid}}{E_l} \times 100, \tag{33}$$

where $E_l = \sum_{j=1}^n E_{l,j}$, is the total amount of energy consumed in the house, and $E_{grid} = \left(\sum_{j=1}^n P_{G/ES,j} + P_{G/L,j} \right) \times t$ is the total amount of imported energy from the main electricity grid.

3.4. Techno-economic model

In this section, the techno-economic model for this platform is delineated. Table 6 lists the considered financial and certain technical parameters that impact the economic performance of the model. It is worth to add, forecasting economic parameters, such as inflation rates, and future cost trends is a challenging task. These forecasts can drastically diverge from actual values due to several unpredicted events. However, estimating the payback time, without considering economic factors, may lead to crucially unrealistic results. Payback time is the duration required to recover an investment. The primary factors in calculating the payback time are the initial installation cost of the setup, inflation, and the economic performance of the system during operation.

There are two methods to calculate this parameter: simple and dis-

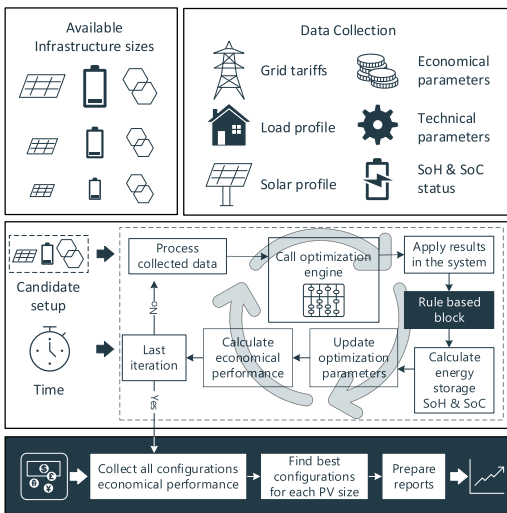


Fig. 5. Proposed platform for PV-BESS size optimization and techno-economic analysis of the deployed setup.

Table 6
Financial and technical parameters.

Parameter	Value	Unit	Symbol
Li-ion BESS cost	750	€/kWh	U_{BESS}
BESS annual price reduction [8]	12%	-	Ω_{BESS}
BESS value/size reduction	3%	Per kWh	$\Omega_{BESS/size}$
PV cost [21]	1300	€/kW	U_{PV}
PV value/size reduction	4%	Per kW	$\Omega_{PV/size}$
PV lifetime [29]	30	Year	Ψ_{PV}
ER cost [69]	250	€/kW	U_{ER}
ER lifetime [29]	15	Year	Ψ_{ER}
ER efficiency [69]	98	%	Ω_{INF}
EMU cost	200	€/unit	U_{EMU}
Load annual increase rate	3%	Per year	Ω_L
Utility tariffs inflation rate	5%	Per year	Ω_{UT}
Annual inflation rate	4%	Per year	Ω_{AN}
Interest rate	3.5%	Per year	Ω_{IN}
Feed-in tariff coefficient	0.8	-	Υ
Maintenance cost [70]	1% of setup cost	€/year	U_M

counted. The annual inflation rate is a crucial factor in computing returns, rendering the simple payback calculation method imprecise due to its exclusion of this rate. Consequently, the discounted payback time ϑ , which incorporates inflation rates in the return calculation, is more accurate and is computed as:

$$\vartheta = \ln \left(\frac{1}{1 - \frac{\xi_c \Omega_{AN}}{\pi_Y}} \right) \div \ln(1 + \Omega_{AN}), \quad (34)$$

where π_Y represents the total annual cash flow in the system, ξ_c denotes the initial investment, and Ω_{AN} signifies the annual inflation. However, due to the impact of system fading on performance over time, causing π_Y to vary each operational year, modifications to the conventional discounted payback formula are necessary to yield accurate results. Consequently, Eq. (33) is modified as:

$$\vartheta_u = -\xi_c - \xi_A + \sum_{k=1}^m \frac{\varphi_k - \Phi_k}{(1 + \frac{\Omega_{AN}}{12})^k}, \quad (35)$$

where ϑ_u is the remaining investment amount for compensation, ξ_A is the annual operating cost of the system, and φ_k is the utility bill in the month number k if the user has not installed any solar or energy storage. For instance, considering a corresponding k_{th} month to be 30 days, $\varphi_k = \sum_{j=1}^{24 \times 30} E_{L,j} \times U_{ToU,j}$. In Eq. (35), Φ_k represents monthly net utility bill after deploying the PV-BESS setup and calculated as follows:

$$\Phi_k = \sum_{j=1}^{24 \times 30} [(E_{G,j} - \Psi E_{f,j}) U_{ToU,j} + E_{Ex,j} C_{Ex}] + C_{f,k}. \quad (36)$$

In Eq. (35), the term $\sum_{k=1}^m \frac{\varphi_k - \Phi_k}{(1 + \frac{\Omega_{AN}}{12})^k} - \xi_A$, represents the net present value of the profit the user accrues after equipping the house with the PV-BESS setup. ξ_c is compiled as:

$$\xi_c = U_{BESS} C_{nm} (1 - \Omega_{BESS/size} C_{nm}) + U_{PV} \Theta_{PV} (1 - \Omega_{PV/size} \Theta_{PV}) + U_{ER} ER_{size} + U_{EMU}, \quad (37)$$

where Θ_{PV} is the solar system size (kWp), U_s is the unit price of installing each of the PV, BESS, ER, and EMU. $\Omega_{s/size}$

Algorithm 1. Economical parameters calculation

(continued on next column)

(continued)

-
- 1: Input: $\vec{E}_g, \vec{E}_f, \vec{U}_{ToU}, \vec{PV}_{size}, \vec{BESS}_{size}, \vec{ER}_{size}$, and Table 6 parameters.
 - 2: **for** v, s, r **in** $\vec{PV}_{size}, \vec{BESS}_{size}, \vec{ER}_{size}$ **do**
 - 3: Calculate $\xi_c[v, s, r]$ based on Eq. (37).
 - 4: Calculate $\xi_A[v, s, r]$ based on Eq. (38).
 - 5: **for** k **in** m **do**
Calculate $\Phi_k[v, s, r]$ based on Eq. (36):
 - 6: $\Phi_k \leftarrow \sum_{j=1}^{24 \times 30} [\vec{E}_{g,v,s,r,j} - \Psi E_{f,v,s,r,j}] U_{ToU,j} + E_{Ex,v,s,r,j} C_{Ex} + C_{f,x}$
Calculate $\vartheta_{u,k}[v, s, r]$ based on Eq.(35):
 - 7: $\vartheta_{u,k} \leftarrow -\xi_c[v, s, r] - \xi_A[v, s, r] + \sum_k \Phi_k$
 - 8: **if** $\vartheta_{u,k} \geq 0$ **then**
 - 9: payback[v, s, r] = k
 - 10: **end if**
 - 11: **end for**
 - 12: Calculate $\delta[v, s, r]$ based on Eq. (32)
 - 13: Calculate $\varepsilon[v, s, r]$ based on Eq. (33)
 - 14: **end for**
 - 15: **for** v **in** \vec{PV}_{size} **do**
 - 16: Select $s' \in s, r' \in r$ which satisfies:
 $\forall (s, r), \vartheta_{u,k}[v, s', r'] \leq \vartheta_{u,k}[v, s, r]$
 - 17: **end for**
 - 18: Return: optimal setup values, $\vec{\vartheta}, \vec{\delta}, \vec{\varepsilon}$
-

represents the cost/size reduction rate for both energy storage and PV panels. Also, ξ_A is compiled as:

$$\xi_A = \xi_c U_M (1 + \Omega_{An})^y + \kappa_{BESS} U_{BESS} (1 + \Omega_{An} - \Omega_{BESS})^{\Psi_{BESS}} + \kappa_{ER} U_{ER} (1 + \Omega_{An})^{\Psi_{ER}}, \quad (38)$$

where y represents the running age of the setup in years, and $U_M \in [0, 1]$ represents the annual maintenance cost ratio, and here is assumed to be 1%. κ_{ER} and κ_{BESS} denote the number of times the ER and energy storage have been replaced respectively during setup run, Ψ_s indicates the year in which the replacement occurred, and Ω_{BESS} represents the annual price reduction in the market for the energy storage technology. Based on Eq. (35), the payback time is the first month in which the ϑ_u turns to a positive value.

In Fig. 5, the box which has been highlighted with darker color, dedicates to economic factors calculation procedure. Algorithm 1 summarizes the calculation procedure in the highlighted section.

4. Results and discussions

In this section, the performance of the proposed system, applied in the mentioned case study, will be demonstrated, and discussed in detail. All simulations and calculations are compiled with a desktop¹ running MATLAB 2022 and Python 3.10. Fig. 6 shows the demand request management in the house for three different setup configurations. One can observe that the purchased amount of energy from main grids increased dramatically when PV-BESS size is limited. Furthermore, providing a proper PV and energy storage sizes gives the EMU the opportunity to be less dependable to the grid. Moreover, Table 7 summarizes the total imported and exported energy from the utility grid, the total stored energy in the BESS, the total solar energy generation, and demand response statuses from the utility grid, BESS, and PV, as well as

¹ Intel(R) Core(TM) i7-7700 K CPU @ 4.20 GHz, 32 GB RAM.

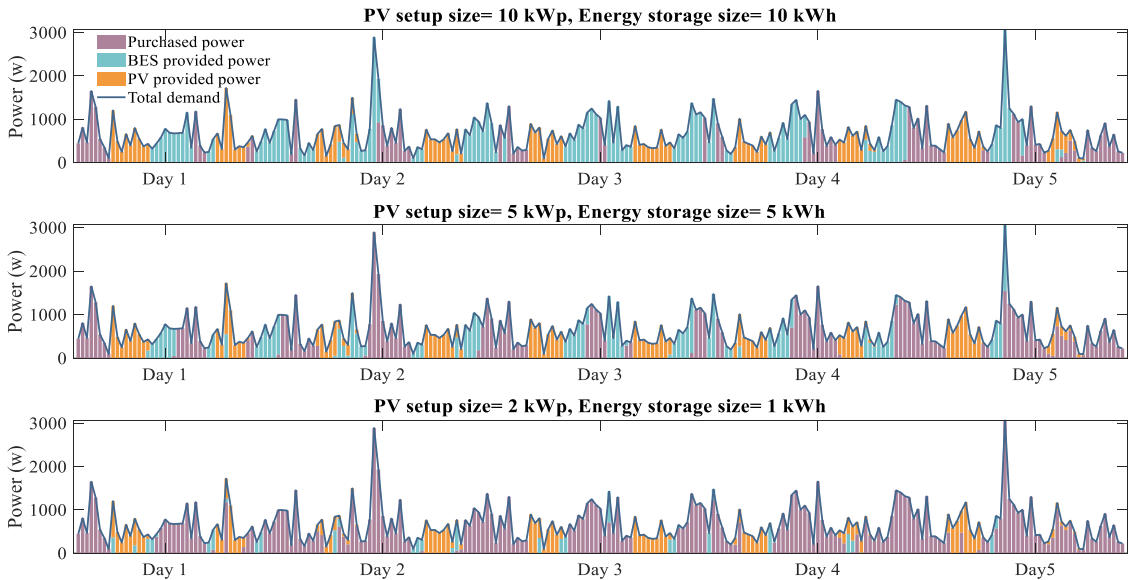


Fig. 6. Demand response management with three different configurations.

Table 7
Different configurations energy status reports during test period (20 years).

Case	1*	2*	3*
Name	Value (kWh)	Value (kWh)	Value (kWh)
Total Export Energy	1.76E5	7.64E4	4.73E3
Total Import Energy	1.06E5	9.84E4	1.02E5
Total Energy Stored in BESS	1.48E5	9.82E4	1.78E4
Total Harvested Solar Energy	1.88E5	9.41E4	1.88E4
Total Demand Response with BESS	6.46E4	6.05E4	1.63E4
Total Demand Response with Utility	0.83E4	3.19E4	8.72E4
Total Demand Response with Solar Energy	4.32E4	2.37E4	1.26E4
Total Demand	1.16E5	1.16E5	1.16E5
Total Energy Waste	2.90E4	1.69E4	1.42E3

*1: PV size = 10 kWp, BESS size = 10 kWh, ER = 5 kW *2: PV size = 5 kWp, BESS size = 5 kWh, ER = 5 kW, *3: PV size = 2 kWp, BESS size = 1 kWh, ER = 2.5 kW.

total energy waste during the simulation period (20 years). Note that the demand profile is the same for all setups, since only one house was selected for investigation and optimization.

In order to investigate the system performance in more detail, the optimization engine performance, the power flow, demand response, and energy storage operation inside the residential grid has been reported for two various conditions in different days. The first scenario has been selected when there is no power generation from solar setup, and system has to respond the demand relying only on the main utility grid and BESS. The second scenario has been picked from samples in which the solar power is enough for demand response, and the optimization engine has the opportunity to decide if the surplus power should be stored or exported to the utility grid to earn profit. Fig. 7(a), and Fig. 7(b) show both scenarios, respectively. In Fig. 7, the power flow between two points is denoted using the 'X to Y' notation, where 'X' represents the energy source and 'Y' the energy consumer. For example, 'PV to Grid' signifies power being exported from the PV system to the grid.

By focusing on power flow trends in both scenarios, it is noticeable that when price fluctuations are high, the charging cycles of the ESS—indicated by a green trajectory in the 'ESS Operation Logs'

subfigure of Fig. 7—primarily occur during periods when energy prices are at their lowest, especially during nighttime. Conversely, the discharging process, marked by a red trajectory in the same subfigure, aligns with periods of high tariffs, typically in the latter half of the day. This trend enables the EMU to benefit from energy arbitrage, thereby minimizing the energy bill. Specifically, in Fig. 7(a), due to the absence of solar power generation, the only options for savings are through exploiting the possibility of energy arbitrage in both demand response and energy trade. Consequently, the EMU sends the charge command to the BESS before the morning peak hours and later discharges the BESS when the energy price rises, starting from 17:00 and ending at 20:00.

During periods of excess solar generation and after demand response actions by renewable energy sources, the EMU has two options for managing surplus energy: storing it in the ESS or exporting it to the grid. In Fig. 7(b), the surplus green energy is used to both charge the BESS and export to the grid. Later, during peak hours when solar energy is not available, the system utilizes the stored energy to meet load demands and engages in energy trading with the grid to earn profit. Additionally, when the EMU anticipates the next day's ideal energy generation forecasts, it can capitalize on the energy trading option even during midnight hours when energy prices are high. This strategy is evident in the early hours depicted in Fig. 7(b).

Energy storage capacity degradation significantly affects the system's performance in terms of both revenue acquisition and the quality of demand response. In all analyses, it is assumed that all the cells are fresh in the initial point. The pivotal role of BESS nominal size in influencing the rate of battery aging is evident from the simulation results. Fig. 8 elucidates the SoH level of energy storage for all simulated configurations, with Fig. 8(a) illustrating a more gradual degradation in larger battery packs compared to smaller ones. This can be attributed to the slower increase in the number of charge and discharge cycles experienced by larger battery packs during operation. Nonetheless, the aging rate of the batteries are multifactorial, being affected not only by the cycle number but also by other elements such as cell temperature, DoD level, and cell current during charges and discharges. In practical experiments, it has been recorded that capacity fading exhibits an exponential trend for Li-ion battery cells [64]. Generally, assuming normal operating conditions—encompassing one charge/discharge

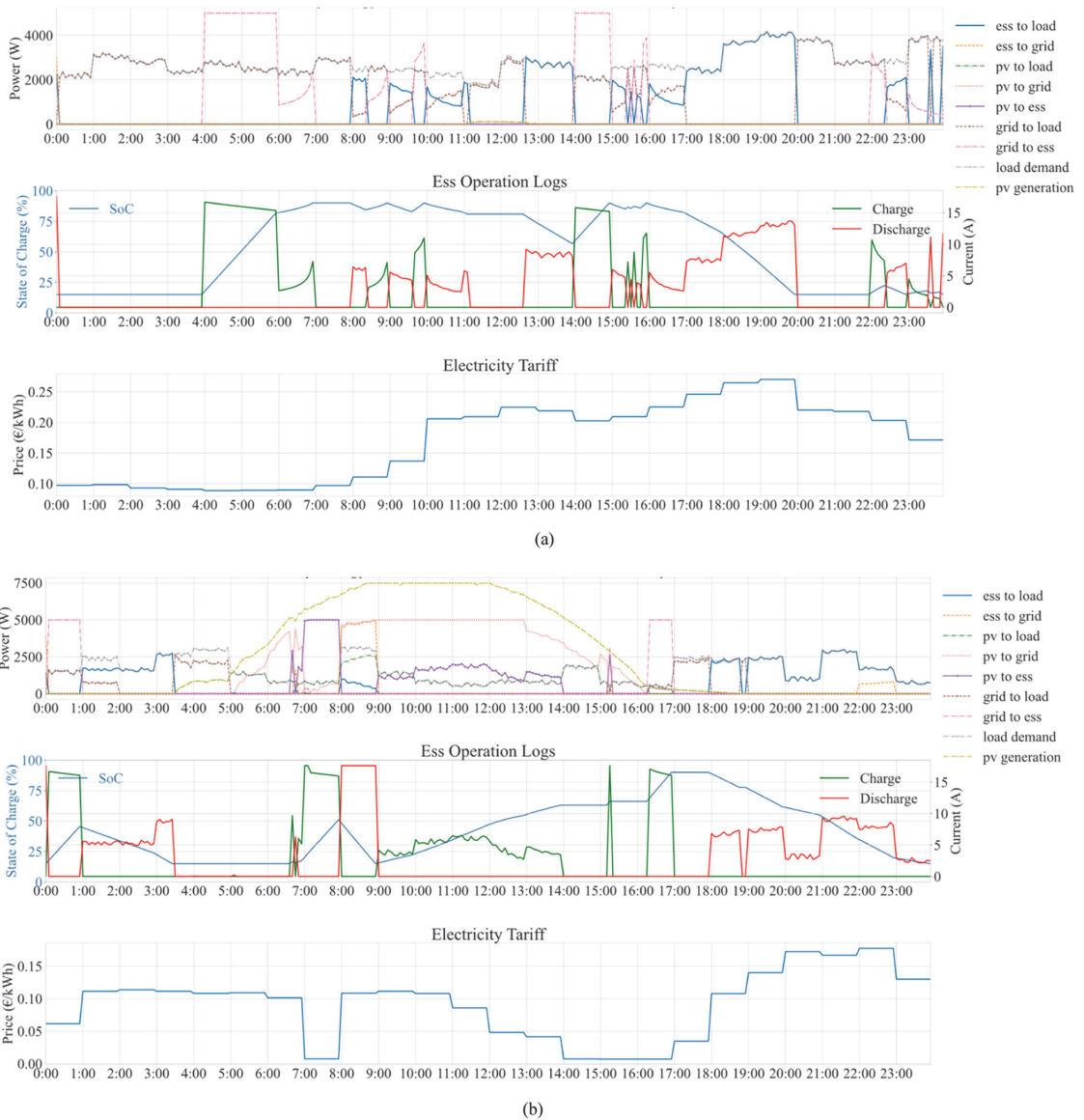


Fig. 7. Daily power flows inside the residential distribution grid. (a) When solar power generation is very limited. (b) When the solar power generation surpluses the demand. Setup configuration: PV size = 7.5 kWp, BESS size = 15 kWh, ER size = 5 kW.

cycle per day, standard operating temperature below 40 °C, and DoD—the linear annual degradation rate for Li-ion batteries can be estimated to range between 2 % and 4 %.

Moreover, the PV size has a bearing on the battery packs’ degradation rate; larger solar power generation can lead to increased battery usage and, consequently, additional degradation. This correlation is demonstrated in Fig. 8(b). Comparing Fig. 8(a) and Fig. 8(b) reveals that the SoH level is predominantly influenced by battery size over PV power capacity. This observation is attributed to the fact that setups with larger solar power generation units tend to export excess energy to the main grid more, thus earning revenue, reducing payback time.

Fig. 9 shows investment returns (dotted red line) and other pertinent

economic factors for the studied house, illustrating the annual cash flow, capital, maintenance, and replacement costs for a case study house in Estonia. In the initial decade, the owner is primarily responsible for annual maintenance and capital costs. However, in the second decade, replacement costs for BESS and power electronic devices are incurred, typically arising at the 10th and 15th years of operation, respectively. This fee is represented as a step down in the cash revenue graphs, aligning with the replacement costs. In all configurations, if any BESS setup reaches its EoL point before the expiration of its warranted lifetime, it is promptly substituted with a new battery pack.

Fig. 10(a) presents a discounted payback time analysis for all PV-BESS combinations. Given the four-dimensional nature of the

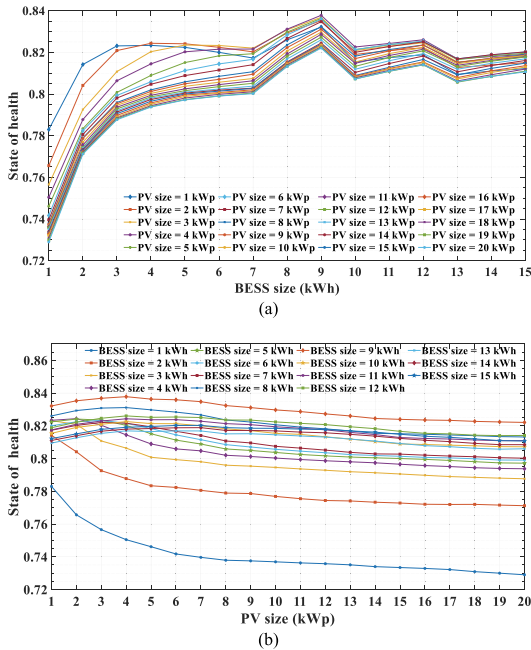


Fig. 8. Battery degradation ratio for all possible combinations of PV-BESS. (a) Degradation level dependency on BESS size. (b) Degradation level dependency on solar generation setup size.

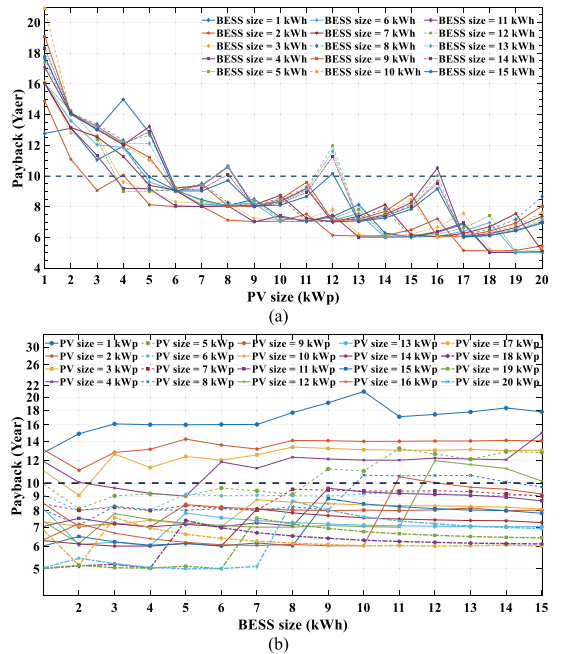


Fig. 10. Discounted payback calculation of PV-BESS combinations. (a) For each BESS size when the PV size is swept in its range. (b) For each PV size when the BESS size is swept in its range. The Y axis has a logarithmic scale in (b).

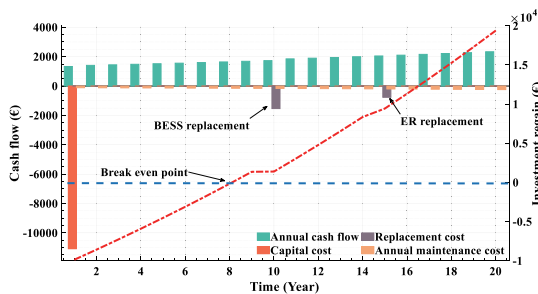


Fig. 9. Economic performance analysis of the case study house in Estonia. Setup configuration: PV size = 5 kWp, BESS size = 5 kWh, ER size = 5 kW.

generated results, the ER dependency is omitted to facilitate clearer visualization, and only the most suitable ER is selected for each PV-BESS size. From the outcomes, the minimum achievable payback time is observed to be around 5 years. The blue horizontal dashed line in Fig. 10 (a) and Fig. 10(b) illustrates the profitability frontier. Here, 10 years is specified in accordance with the lifespan of battery packs. Thus, if the system cannot reach the break-even point within the aforementioned time, it is deemed economically impractical. The results reveal that considering the energy market condition in 2022, choosing the size of PV systems larger than 9 kWp will mostly yield income for the owner.

Fig. 10(b) indicates the dependencies of payback time on BESS size for each specific PV system. This observation demonstrates that, in Estonia, for residential applications, PV-BESS systems with a PV size less than 4.5 kWp are not profitable, even with very small-sized BESSs. Moreover, the impact of PV size on payback time is substantially more pronounced than the impact of BESS size. This is predominantly

because, with partial green energy generation, the majority of BESS capacity is allocated to energy arbitrage. Nevertheless, due to battery protection policies and constraints in power electronics, energy arbitrage exhibits lesser dependence on BESS size variations when the PV size is held constant. Consequently, in the present circumstances in Estonia, solely installing BESS in houses without PV systems does not emerge as a feasible solution for long-term use.

Table 8 displays the most profitable BESS-ER sizes for each PV arrangement, along with the associated repayment periods and expected income after 20 years of operation. These outcomes are derived from evaluating two distinct scenarios. The first scenario takes into account the current incentive under the feed-in policy, while the second scenario assumes the termination of the feed-in policy for residential applications. In the latter scenario, surplus solar energy is channeled to the BESS, optimizing system benefits to meet load demands; thus, no electricity is fed back into the utility grid.

The findings in Table 8 serve as practical guidelines for end-users regarding the anticipated payback periods of varied setups in Estonia. However, it is crucial to acknowledge that these results are deduced by analyzing a specific load profile, implying that alterations in load profiles would yield different outcomes. Therefore, while these results provide a valuable point of reference, adjustments may be necessary to accurately reflect the payback periods and profitability of different setups based on the individual load profiles and other specific circumstances pertinent to each case.

The extent of achieved energy autonomy is illustrated in Fig. 11(a) and Fig. 11(b). In scenarios utilizing energy trade options, the maximum attained energy autonomy is approximately 30%. In Fig. 11(a), red segments on the plot represent areas of negative energy autonomy, signifying that energy imported from the main grid surpasses total load demand. This phenomenon predominantly occurs in setups with expansive storage sizes. Typically, in buildings utilizing PV-BESS, a boost in storage size is presumed to elevate energy autonomy. This

Table 8
Most profitable combinations' economical results.

PV size (kWp)	FIT	BESS size (kWh)	ER size (kW)	Payback (year)	Revenue (€)
1	0.8 ToU	1	2.5	12.25	2484
-	-	1	2.5	17.1	507
2	0.8 ToU	2	2.5	11.09	5603
-	-	2	2.5	18.62	841
3	0.8 ToU	2	2.5	9.05	8524
-	-	2	2.5	19.15	349
4	0.8 ToU	3	5	9.6	11463
-	-	2	5	19.63	106
5	0.8 ToU	4	5	9.1	14528
-	-	2	5	> 20	-1097
6	0.8 ToU	4	5	8.06	17779
-	-	2	5	> 20	-1822
7	0.8 ToU	5	7.5	8.4	21117
-	-	3	7.5	> 20	-2480
8	0.8 ToU	5	7.5	8.03	24556
-	-	3	7.5	> 20	-3095
9	0.8 ToU	6	10	8.22	27983
-	-	3	10	> 20	-3653
10	0.8 ToU	6	10	7.65	31516
-	-	3	10	> 20	-4159
11	0.8 ToU	7	10	7.25	35117
-	-	3	10	> 20	-4614
12	0.8 ToU	8	12.5	7.01	38574
-	-	4	12.5	> 20	-5014
13	0.8 ToU	9	12.5	7.18	42231
-	-	4	12.5	> 20	-5368
14	0.8 ToU	9	12.5	7.69	45956
-	-	4	12.5	> 20	-5672
15	0.8 ToU	10	15	8.45	49566
-	-	4	15	> 20	-5931
16	0.8 ToU	10	15	6.03	53429
-	-	4	15	> 20	-6146
17	0.8 ToU	12	15	6.02	57269
-	-	4	15	> 20	-6321
18	0.8 ToU	14	17.5	6.11	61139
-	-	5	17.5	> 20	-6456
19	0.8 ToU	15	17.5	6.42	65154
-	-	5	17.5	> 20	-6559
20	0.8 ToU	15	20	6.93	69126
-	-	5	20	> 20	-6631

unexpected observation can be attributed to the opportunities for energy arbitrage. Hence, elevated FiTs and the profitable nature of energy arbitrage make setups with substantial energy storages prefer energy trade, engaging in more energy exchange than the actual load necessities. Consequently, in setups where energy trade is an option, solely considering the energy autonomy factor as a metric understates a system's performance.

In contrast, Fig. 11(b) displays the opposite trend due to restrictions on delivering energy to the main grid. Thus, the supervisory algorithm reserves all surplus PV-generated electricity in the energy storage for responding to future load demands, diminishing dependency on the main grid, and thereby increasing the overall energy autonomy ratio. As depicted in the surface plot, augmenting BESS size improves the building's energy autonomy. In this scenario, the energy autonomy factor relies more heavily on the PV system size than the BESS size. In this contrasting scenario, the maximum achievable energy autonomy factor reaches up to 48%. In Tallinn, the long, dark winter days make achieving high energy autonomy a significant challenge. In some instances, meeting demand without the main grid's support during consecutive cloudy days becomes nearly impossible.

The PV self-consumption factor for one selected BESS size and range of PV setups, considering both feed-in possibilities and energy exchange restrictions, is depicted in Fig. 12. This metric exhibits higher values when the algorithm lacks a trade option and consumes as much energy as the PV system generates. However, when a trade option is available, the supervisory system opts to exchange excess available energy with the main grid and utilizes stored off-peak energy to satisfy load demand.

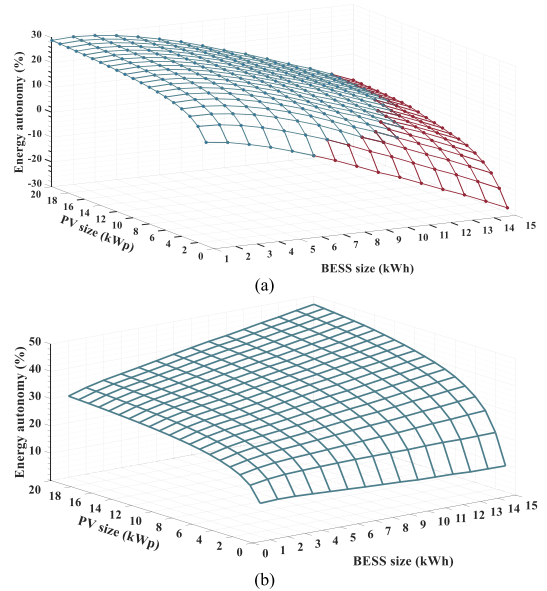


Fig. 11. Energy autonomy factor for different PV-BESS sizes. (a) Considering feed-in possibility. (b) Without feed-in possibility.

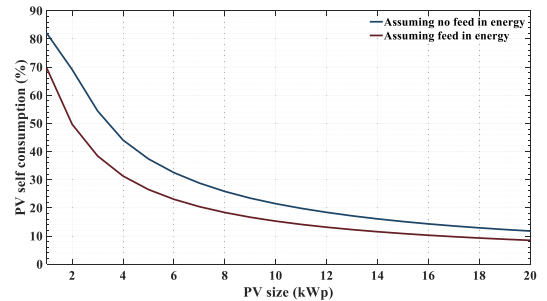


Fig. 12. PV self-consumption ratios comparison between two scenarios.

This approach reduces the PV self-consumption ratio but increases revenue. Moreover, as the size of the PV system expands, the amount of green energy generated exceeds the load demand, leading to the majority of produced energy being delivered to the main grid or remaining unused. Thus, achieving high PV self-consumption ratios is a challenge in Estonia even with substantial PV-BESS systems, as well. However, cooperative policies for clean energy production and daily intensive tariff fluctuations have forged significant opportunities for energy arbitrage.

In Table 9, state-of-the-art results are compiled to facilitate an analysis of PV-BESS installation payback times across diverse locations. Notably, the proposed algorithm achieved a lower payback period in comparison to state-of-the-art. However, due to the distinct and specific variables and parameters inherent to each system, drawing straightforward comparisons among the literature is not simple. Another significant variable affecting payback period calculations encompasses the extensive range of economic parameters in different studies. For instance, the capital costs for BESS and PV systems show considerable variability, with prices spanning from €100 to €1200 and €400 to €1550, respectively. Furthermore, factors such as climatic conditions, the

Table 9
Proposed model benchmarking with state-of-the-art.

Ref.	Location	Currency	UTs/kWh	FiTs/kWh	Tariff type	U _{PV} /kWp	U _{BESS} /kWh	U _M	BESS fade	Inflation	δ (Years)
PP	Estonia	Euro	Nord pool	0.8*ToU	ToU	1300	750	✓	✓	✓	5
†[11]	Australia	AUD	×	×	ToU	400	740	✓	✓	✓	9
†[29]	South Africa	USD	0.11	0.16	FT	651	405	✓	×	✓	8.6
†[12]	Estonia	Euro	Nord pool	×	FT/SP/ToU	400	100	×	×	×	10
†[6]	Italy	Euro	0.19	0.06	FT	700	500	✓	×	✓	Profitable
†[13]	Switzerland	USD	0.17–0.25	0.056	FT/DT/ToU	0	450	×	×	✓	Unprofitable
†[19]	Taiwan	USD	0.04–0.15	×	ToU	0	340	✓	×	✓	8.37
†[21]	Netherlands	USD	0.20	0.092	ToU	1350	500	✓	✓	×	11
†[21]	USA	USD	0.11	0.097	ToU	1350	500	✓	✓	×	20
†[15]	Ireland	Pound	0.08–0.32	0.05	FT/DT/TT	1400	507	✓	×	✓	22
†[4]	Germany	Euro	0.443	0.108	FT	1542	1178	✓	✓	✓	Profitable
†[22]	Australia	USD	0.107–0.511	0.06	ToU	1500	148	×	×	×	5–10

geographical locale of the installation, existing incentive policies, can alter the strategy of the control algorithm, leading to diverse outcomes.

Due to the intermittent nature of renewable resources and uncertainties related to energy consumption profiles within dwellings, optimizing energy flow without accounting for these natural uncertainties will lead to deviations from ideal operational assumptions. To analyze the effects of uncertainties on the system’s operational logic, Monte Carlo simulation has been performed. To this end, random errors have been introduced into the PV and load profiles. These errors are assumed to follow a Gaussian distribution with a mean value of zero and standard deviations of 0.05 and 0.2, corresponding to error levels of 5 % and 20 %, respectively. Fig. 13(a) and Fig. 13(b) display the error injections to the daily PV and load profiles.

The outcomes of these random error injections have been analyzed after 100 iterations. Deviations in the system’s variables are depicted in Fig. 14(a)–(d). It is observed that the parameter most affected among the demand response elements is the power flow from energy storage to the load. Conversely, the energy storage’s SoC profile shows the least deviation from ideal performance, despite these variations. This phenomenon can be attributed to the fact that the deviations do not persist for long periods, and the impact of errors on energy flow optimization has a short-lived effect. Yet, at certain times, the EMU’s decisions regarding energy flow supervision and demand response follow entirely different scenarios, especially in cases of 20% error injection. This leads the algorithm to converge to different minima, indicating that the system’s robustness should be examined more thoroughly in future studies.

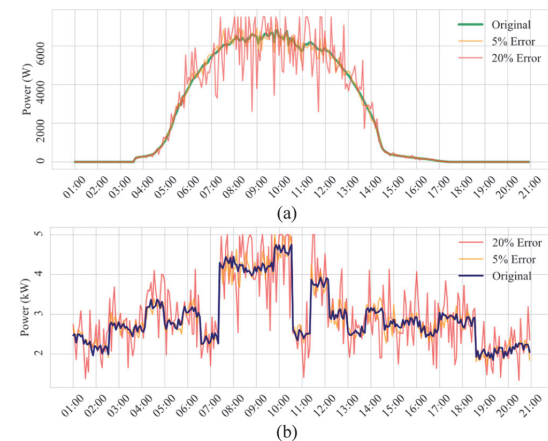


Fig. 13. Error injections to PV and load profiles. (a) PV profile fluctuations with error injections. (b) Load profile fluctuations with error injections.

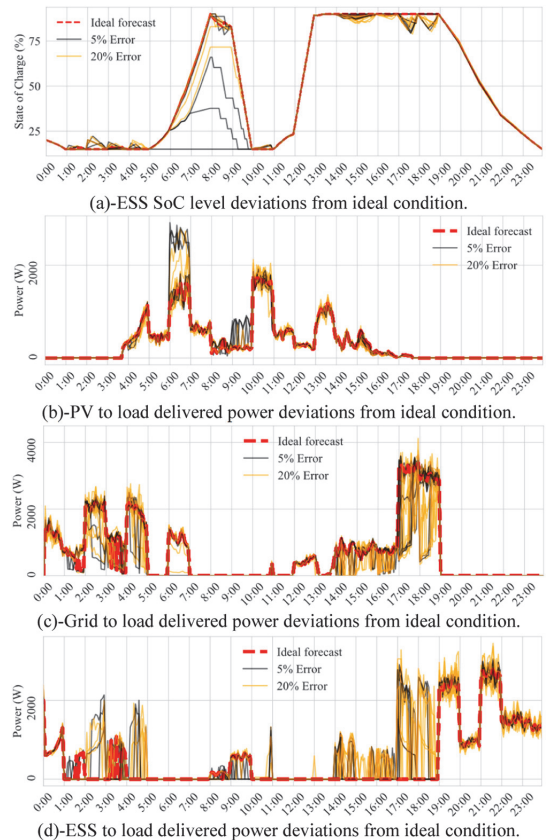


Fig. 14. System variables deviations from ideal condition in presence of error injection to PV and Load profiles.

5. Conclusion

This paper provides a comprehensive analysis of the economic viability of residential PV-BESS setups in the Estonian energy market, taking into account a multitude of technical and economic parameters. A battery degradation model is incorporated in this study to accurately simulate battery capacity loss during energy processes, considering both calendar aging and cyclic degradation. The personalized techno-economic model integrates various economic variables and parameters including annual interest rates, price size dependencies, net present

value, maintenance and replacement costs, and government policy incentives.

Additionally, a novel rule-based non-linear optimization function is implemented to maximize system revenue while minimizing energy storage degradation. The rules within the system are formulated to maximize benefits from energy arbitrage and FIT incentives, in accordance with battery aging constraints. Furthermore, by analyzing the outcomes of Monte Carlo simulation, the robustness of the optimization framework against certain levels of uncertainties in input parameters has been assessed. Such investigations are crucial to enhancing the system's reliability and robustness.

The findings of this paper indicate that installing hybrid PV-BESS systems in Estonia can currently be profitable, with minimum payback time of 5 years, but this profitability is significantly dependent on FITs incentives and ToU fluctuations. However, the removal of FITs incentives notably impacts the profitability of the systems, making hybrid PV-BESS setups economically non-viable without such subsidies. In such a scenario, owners could only rely on energy self-sufficiency and grid relief. Furthermore, the maximum achievable energy autonomy ratios in investigated scenarios are 30% and 48% with and without a feed-in option, respectively, which shows the unfeasibility of grid independence. Despite the current potential for profitability, the evolving economic landscape necessitates a careful consideration of the long-term economic outlook for PV-BESS installations in Estonia.

The main goals and contributions of this work are as follows:

- Providing a precise and reliable personalized tool for end-users and authorities to customize the calculations for obtaining the optimal size of PV-BESS and inverter configuration based on their load profile, energy generation, and electricity tariff profile. This tool will support them, in the decision-making process, by delivering capital cost, best-matching instrument combinations, and budget-based payback time estimation. Furthermore, the model's applicability is not confined to any specific region; it can be adapted and utilized for personalized scenarios across diverse geographical locations.
- The effect of element fading is included in profit maximization equations and discounted payback calculations. Thus, the conventional discounted payback formula is modified to enhance payback time result accuracy.
- An optimization engine designed by fusion of rule-based, non-linear gradient based, and grid search methods, to benefit each methods strength for channeling energy inside the residential grid.

CRedit authorship contribution statement

Oleksandr Husev: Conceptualization, Project administration, Supervision, Validation, Writing – review & editing. **Hossein Nourollahi Hokmabad:** Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Jarek Kurnitski:** Conceptualization, Project administration, Supervision, Validation, Writing – review & editing. **Juri Belikov:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgment

This work was supported by the Estonian Research Council grants PRG675, PRG1463, and Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230) funded by the Estonian Ministry of Education and Research. We confirm that the manuscript has been read and approved by all named authors and that there are no other people who satisfied the criteria for authorship but are not listed.

References

- [1] E. Tsioumas, N. Jabbour, M. Koseoglou, C. Mademlis, A novel control strategy for improving the performance of a nearly zero energy building, *IEEE Trans. Power Electron.* vol. 35 (2) (2019) 1513–1524.
- [2] M. Elkazaz, M. Sumner, E. Naghiyev, Z. Hua, D.W. Thomas, Techno-Economic Sizing of a community battery to provide community energy billing and additional ancillary services, *Sustain. Energy, Grids Netw.* vol. 26 (2021) 100439.
- [3] L. Sang, Y. Xu, H. Long, Q. Hu, H. Sun, Electricity price prediction for energy storage system arbitrage: A decision-focused approach, *IEEE Trans. Smart Grid* vol. 13 (4) (2022) 2822–2832.
- [4] A. Dietrich, C. Weber, What drives profitability of grid-connected residential PV storage systems? A closer look with focus on Germany, *Energy Econ.* vol. 74 (2018) 399–416.
- [5] A.A.R. Mohamed, R.J. Best, X. Liu, D.J. Morrow, A comprehensive robust techno-economic analysis and sizing tool for the small-scale PV and BESS, *IEEE Trans. Energy Convers.* vol. 37 (1) (2021) 560–572.
- [6] I. D'Adamo, M. Gastaldi, P. Morone, The impact of a subsidized tax deduction on residential solar photovoltaic-battery energy storage systems, *Util. Policy* vol. 75 (2022) 101358.
- [7] Angelos I. Nousedis, K. Kryonidis Georgios, Eleftherios O. Kontis, Georgios A. Barzegkar-Ntovom, Ioannis P. Panapakidis, Georgios C. Christoforidis, Grigoris K. Papagiannis, Impact of policy incentives on the promotion of integrated PV and battery storage systems: a techno-economic assessment, *IET Renew. Power Gener.* vol. 14 (7) (2020) 1174–1183.
- [8] D. Gardiner, O. Schmidt, P. Heptonstall, R. Gross, I. Staffell, Quantifying the impact of policy on the investment case for residential electricity storage in the UK, *J. Energy Storage* vol. 27 (2020) 101140.
- [9] M.N. Ashtiani, A. Toopshekan, F.R. Astaraei, H. Yousefi, A. Maleki, Techno-economic analysis of a grid-connected PV/battery system using the teaching-learning-based optimization algorithm, *Sol. Energy* vol. 203 (2020) 69–82.
- [10] M.M. Symeonidou, C. Zioga, A.M. Papadopoulos, Life cycle cost optimization analysis of battery storage system for residential photovoltaic panels, *J. Clean. Prod.* vol. 309 (2021) 127234.
- [11] J. Every, L. Li, D.G. Dorrell, Optimal selection of small-scale hybrid PV-battery systems to maximize economic benefit based on temporal load data, *12th IEEE Conf. Ind. Electron. Appl. (ICIEA)* (2017) 471–476.
- [12] N. Shabbir, L. Kütt, V. Astapov, M. Jawad, A. Allik, O. Husev, Battery size optimization with customer PV installations and domestic load profile, *IEEE Access* vol. 10 (2022) 13012–13025.
- [13] A. Pena-Bello, M. Burer, M.K. Patel, D. Parra, Optimizing PV and grid charging in combined applications to improve the profitability of residential batteries, *J. Energy Storage* vol. 13 (2017) 58–72.
- [14] E. Doroudchi, S.K. Pal, M. Lehtonen, J. Kyyrä, Optimizing energy cost via battery sizing in residential PV/battery systems, *IEEE Innov. Smart Grid Technol. -ASIA (ISGT ASIA)* (2015) 1–6.
- [15] R. Machlev, N. Zargari, N. Chowdhury, J. Belikov, Y. Levron, A review of optimal control methods for energy storage systems-energy trading, energy balancing and electric vehicles, *J. Energy Storage* vol. 32 (2020) 101787.
- [16] M.A. Hannan, M. Faisal, P.J. Ker, L.H. Mun, K. Parvin, T.M.I. Mahlia, F. Blaabjerg, "A review of internet of energy based building energy management systems: issues and recommendations," *IEEE Access* vol. 6 (2018) 38997–39014.
- [17] B. Gu, C. Mao, B. Liu, D. Wang, H. Fan, J. Zhu, Z. Sang, Optimal charge/discharge scheduling for batteries in energy router-based microgrids of prosumers via peer-to-peer trading, *IEEE Trans. Sustain. Energy* vol. 13 (3) (2022) 1315–1328.
- [18] P. Singh, S. Dhundhara, Y.P. Verma, N. Tayal, "Optimal battery utilization for energy management and load scheduling in smart residence under demand response scheme, *Sustain. Energy, Grids Netw.* vol. 26 (2021) 100432.
- [19] C.-T. Tsai, E.M. Ocampo, T.M. Beza, C.-C. Kuo, "Techno-economic and sizing analysis of battery energy storage system for behind-the-meter application," *IEEE Access* vol. 8 (2020) 203734–203746.
- [20] V.V. Murty, A. Kumar, Optimal energy management and techno-economic analysis in microgrid with hybrid renewable energy sources, *J. Mod. Power Syst. Clean. Energy* vol. 8 (5) (2020) 929–940.
- [21] S. Bandyopadhyay, G.R.C. Mouli, Z. Qin, L.R. Elizondo, P. Bauer, Techno-economical model based optimal sizing of PV-battery systems for microgrids, *IEEE Trans. Sustain. Energy* vol. 11 (3) (2019) 1657–1668.
- [22] J. Li, "Optimal sizing of grid-connected photovoltaic battery systems for residential houses in Australia," *Renew. Energy* vol. 136 (2019) 1245–1254.
- [23] H.X. Li, P. Horan, M.B. Luther, T.M. Ahmed, Informed decision making of battery storage for solar-PV homes using smart meter data, *Energy Build.* vol. 198 (2019) 491–502.

- [24] M. Yousefi, A. Hajizadeh, M.N. Soltani, B. Hredzak, Predictive home energy management system with photovoltaic array, heat pump, and plug-in electric vehicle, *IEEE Trans. Ind. Inform.* vol. 17 (1) (2020) 430–440.
- [25] K. Parvin, M.H. Lipu, M.A. Hannan, M.A. Abdullah, K.P. Jern, R.A. Begum, M. Mansur, K.M. Muttaqi, T.I. Mahlia, Z.Y. Dong, Intelligent controllers and optimization algorithms for building energy management towards achieving sustainable development: challenges and prospects, *IEEE Access* vol. 9 (2021) 41577–41602.
- [26] H. Yang, Z. Gong, Y. Ma, L. Wang, B. Dong, Optimal two-stage dispatch method for household PV-BESS integrated generation system under time-of-use electricity price, *Int. J. Electr. Power Energy Syst.* vol. 123 (2020) 106244.
- [27] M. Tostado-Véliz, D. Icaza-Alvarez, F. Jurado, A novel methodology for optimal sizing photovoltaic-battery systems in smart homes considering grid outages and demand response, *Renew. Energy* vol. 170 (2021) 884–896.
- [28] G. Soykan, G. Er, E. Canakoglu, “Optimal sizing of an isolated microgrid with electric vehicles using stochastic programming, *Sustain Energy Grids Netw.* vol. 32 (2022) 100850.
- [29] B.P. Numbi, S.J. Malinga, Optimal energy cost and economic analysis of a residential grid-interactive solar PV system- case of eThekweni municipality in South Africa, *Appl. Energy* vol. 186 (2017) 28–45.
- [30] W. Wei, Z. Wang, F. Liu, M. Shafie-khah, J.P. Catalão, Cost-efficient deployment of storage unit in residential energy systems, *IEEE Trans. Power Syst.* vol. 36 (1) (2020) 525–528.
- [31] H. Shin, J.H. Roh, Framework for sizing of energy storage system supplementing photovoltaic generation in consideration of battery degradation, *IEEE Access* vol. 8 (2020) 60246–60258.
- [32] R. Ahmadihangar, H. Karami, O. Husev, A. Blinov, A. Rosin, A. Jonaitis, M. J. Sanjari, Analytical approach for maximizing self-consumption of nearly zero energy buildings-case study: Baltic region, *Energy* vol. 238 (2022) 121744.
- [33] J. Linssen, P. Stenzel, J. Fleer, Techno-economic analysis of photovoltaic battery systems and the influence of different consumer load profiles, *Appl. Energy* vol. 185 (2017) 2019–2025.
- [34] A. Baniasadi, D. Habibi, W. Al-Saedi, M.A. Masoum, C.K. Das, N. Mousavi, Optimal sizing design and operation of electrical and thermal energy storage systems in smart buildings, *J. Energy Storage* vol. 28 (2020) 101186.
- [35] Y. Al-Wreikat, E.K. Attfield, J.R. Sodre, Model for payback time of using retired electric vehicle batteries in residential energy storage systems, *Energy* vol. 259 (2022) 124975.
- [36] A.C. Duman, H.S. Erden, Ö. Gönül, Ö. Güler, Optimal sizing of PV-BESS units for home energy management system-equipped households considering day-ahead load scheduling for demand response and self-consumption, *Energy Build.* vol. 267 (2022) 112164.
- [37] L. Bhamidi, S. Sivasubramani, Optimal sizing of smart home renewable energy resources and battery under prosumer-based energy management, *IEEE Syst. J.* vol. 15 (1) (2020) 105–113.
- [38] R. Khezri, A. Mahmoudi, D. Whaley, Optimal sizing and comparative analysis of rooftop PV and battery for grid-connected households with all-electric and gas-electricity utility, *Energy* vol. 251 (2022) 123876.
- [39] B.S. Sami, N. Sihem, Z. Bassam, “Design and implementation of an intelligent home energy management system: a realistic autonomous hybrid system using energy storage, *Int. J. Hydrog. Energy* vol. 43 (42) (2018) 19352–19365.
- [40] S. Bahramara, Robust optimization of the flexibility-constrained energy management problem for a smart home with rooftop photovoltaic and an energy storage, *J. Energy Storage* vol. 36 (2021) 102358.
- [41] Y. He, S. Guo, P. Dong, C. Wang, J. Huang, J. Zhou, Techno-economic comparison of different hybrid energy storage systems for off-grid renewable energy applications based on a novel probabilistic reliability index, *Appl. Energy* vol. 328 (2022) 120225.
- [42] A. Bhatt, W. Ongsakul, N. Madhu, Optimal techno-economic feasibility study of net-zero carbon emission microgrid integrating second-life battery energy storage system, *Energy Convers. Manag.* vol. 266 (2022) 115825.
- [43] C. Jamroen, Optimal techno-economic sizing of a standalone floating photovoltaic/battery energy storage system to power an aquaculture aeration and monitoring system, *Sustain. Energy Technol. Assess.* vol. 50 (2022) 101862.
- [44] Y. Wu, Z. Liu, J. Liu, H. Xiao, R. Liu, L. Zhang, Optimal battery capacity of grid-connected PV-battery systems considering battery degradation, *Renew. Energy* vol. 181 (2022) 10–23.
- [45] M.B. Mustafa, P. Keatley, Y. Huang, O. Agbonaye, O.O. Ademulegun, N. Hewitt, Evaluation of a battery energy storage system in hospitals for arbitrage and ancillary services, *J. Energy Storage* vol. 43 (2021) 103183.
- [46] J. Liu, Z. Liu, Y. Wu, X. Chen, H. Xiao, L. Zhang, Impact of climate on photovoltaic battery energy storage system optimization, *Renew. Energy* vol. 191 (2022) 625–638.
- [47] M. Aghamohamadi, A. Mahmoudi, M.H. Haque, Two-stage robust sizing and operation co-optimization for residential PV-battery systems considering the uncertainty of PV generation and load, *IEEE Trans. Ind. Inform.* vol. 17 (2) (2020) 1005–1017.
- [48] T.H. Shahsavari, N.V. Kurdkandi, O. Husev, E. Babaei, M. Sabahi, A. Khoshkbar-Sadigh, D. Vinnikov, A new flying capacitor-based buck-boost converter for dual-purpose applications, *IEEE J. Emerg. Sel. Top. Ind. Electron.* vol. 4 (2) (2023) 447–459.
- [49] U. Mulleriyawage, W. Shen, Optimally sizing of battery energy storage capacity by operational optimization of residential PV-Battery systems: an Australian household case study, *Renew. Energy* vol. 160 (2020) 852–864.
- [50] W. Vermeer, G.R.C. Mouli, P. Bauer, Optimal sizing and control of a PV-EV-BES charging system including primary frequency control and component degradation, *IEEE Open J. Ind. Electron. Soc.* vol. 3 (2022) 236–251.
- [51] A. Saez-de Ibarra, E. Martínez-Laserna, D.-I. Stroe, M. Swierczynski, P. Rodríguez, Sizing study of second life Li-ion batteries for enhancing renewable energy grid integration, *IEEE Trans. Ind. Appl.* vol. 52 (6) (2016) 4999–5008.
- [52] N. Horesh, C. Quinn, H. Wang, R. Zane, M. Ferry, S. Tong, J.C. Quinn, Driving to the future of energy storage: Techno-economic analysis of a novel method to recondition second life electric vehicle batteries, *Appl. Energy* vol. 295 (2021) 117007.
- [53] Z. He, J. Khazaei, J.D. Freihaut, Optimal integration of Vehicle to Building (V2B) and Building to Vehicle (B2V) technologies for commercial buildings, *Sustain Energy Grids Netw.* vol. 32 (2022) 100921.
- [54] J. Zhu, I. Mathews, D. Ren, W. Li, D. Cogswell, B. Xing, T. Sedlatschek, S.N. R. Kantareddy, M. Yi, T. Gao, Y. Xia, End-of-life or second-life options for retired electric vehicle batteries, *Cell Rep. Phys. Sci.* vol. 2 (8) (2021).
- [55] 18650 Lithium-Ion Power Cell datasheet, LithiumWerks, Accessed December 2022, [Online].
- [56] B. Xu, A. Oudalov, A. Ulbig, G. Andersson, D.S. Kirschen, Modeling of lithium-ion battery degradation for cell life assessment, *IEEE Trans. Smart Grid* vol. 9 (2) (2016) 1131–1140.
- [57] K. Schwenk, S. Meisenbacher, B. Briegel, T. Harr, V. Hagenmeyer, R. Mikut, Integrating battery aging in the optimization for bidirectional charging of electric vehicles, *IEEE Trans. Smart Grid* vol. 12 (6) (2021) 5135–5145.
- [58] E. Lobato, L. Sigrist, A. Ortega, A. González, J. Fernández, Battery energy storage integration in wind farms: Economic viability in the Spanish market, *Sustain. Energy, Grids Netw.* vol. 32 (2022) 100854.
- [59] M. Haris, M.N. Hasan, S. Qin, Degradation curve prediction of lithium-ion batteries based on knee point detection algorithm and convolutional neural network, *IEEE Trans. Instrum. Meas.* vol. 71 (2022) 1–10.
- [60] N. Khan, F.U.M. Ullah, A.Ullah Afnan, M.Y. Lee, S.W. Baik, Batteries state of health estimation via efficient neural networks with multiple channel charging profiles, *IEEE Access* vol. 9 (2020) 7797–7813.
- [61] K. Liu, Y. Shang, Q. Ouyang, W.D. Widanage, “A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery, *IEEE Trans. Ind. Electron.* vol. 68 (4) (2020) 3170–3180.
- [62] Y. Bi, Y. Yin, S.-Y. Choe, Online state of health and aging parameter estimation using a physics-based life model with a particle filter, *J. Power Sources* vol. 476 (2020) 228655.
- [63] R. Xiong, L. Li, Z. Li, Q. Yu, H. Mu, An electrochemical model based degradation state identification method of Lithium-ion battery for all-climate electric vehicles application, *Appl. Energy* vol. 219 (2018) 264–275.
- [64] S.N. Motapon, E. Lachance, L.-A. Dessaint, K. Al-Haddad, A generic cycle life model for lithium-ion batteries based on fatigue theory and equivalent cycle counting, *IEEE Open J. Ind. Electron. Soc.* vol. 1 (2020) 207–217.
- [65] L. Lam, P. Bauer, Practical capacity fading model for Li-ion battery cells in electric vehicles, *IEEE Trans. Power Electron.* vol. 28 (12) (2012) 5910–5918.
- [66] S. Amir, M. Gulzar, M.O. Tarar, I.H. Naqvi, N.A. Zaffar, M.G. Pecht, Dynamic equivalent circuit model to estimate state-of-health of lithium-ion batteries, *IEEE Access* vol. 10 (2022) 18279–18288.
- [67] X. Bian, L. Liu, J. Yan, Z. Zou, R. Zhao, An open circuit voltage-based model for state-of-health estimation of lithium-ion batteries: model development and validation, *J. Power Sources* vol. 448 (2020) 227401.
- [68] M. Swierczynski, D.-I. Stroe, A.-I. Stan, R. Teodorescu, S.K. Kær, Lifetime estimation of the nano phosphate LiFePO₄ battery chemistry used in fully electric vehicles, *IEEE Trans. Ind. Appl.* vol. 51 (4) (2015) 3453–3461.
- [69] M. Najafzadeh, N. Strzelecka, O. Husev, I. Roasto, K. Nassereddine, D. Vinnikov, R. Strzelecki, Grid-forming operation of energy-router based on model predictive control with improved dynamic performance, *Energies* vol. 15 (11) (2022) 4010.
- [70] M. Mehrtaf, F. Capitanescu, P.K. Heiselberg, T. Gibon, A. Bertrand, “An enhanced optimal PV and battery sizing model for zero energy buildings considering environmental impacts, *IEEE Trans. Ind. Appl.* vol. 56 (6) (2020) 6846–6856.

Publication II

H. N. Hokmabad, O. Husev, and J. Belikov, 'Day-Ahead Solar Power Forecasting Using LightGBM and Self-Attention Based Encoder-Decoder Networks', *IEEE Transactions on Sustainable Energy*, vol. 16, no. 2, pp. 866–879, 2025.

Day-Ahead Solar Power Forecasting Using LightGBM and Self-Attention Based Encoder-Decoder Networks

Hossein Nourollahi Hokmabad¹, *Student Member, IEEE*, Oleksandr Husev¹, *Senior Member, IEEE*, and Juri Belikov¹, *Senior Member, IEEE*

Abstract—The burgeoning trend of integrating renewable energy harvesters into the grid introduces critical issues for its reliability and stability. These issues arise from the stochastic and intermittent nature of renewable energy sources. Data-driven forecasting tools are indispensable in mitigating these challenges with their rugged performance. However, tools relying solely on data-driven methods often underperform when an adequate amount of recorded data is unattainable. To bridge this gap, this paper presents a novel day-ahead hybrid forecasting framework for photovoltaic applications. This framework integrates a physics-based model with Machine Learning (ML) techniques, enhancing prediction reliability in environments with scarce data. Additionally, an innovative ML pipeline is introduced for data-abundant environments. The proposed ML tool comprises two branches: a set of regressors, each tailored for specific weather conditions, and a self-attention-based encoder-decoder network. By fusing the outputs from these branches through a meta-learner, the tool achieves predictions of higher quality, as evidenced by its superior performance over benchmark models in an investigated test dataset.

Index Terms—Solar power forecasting, photo voltaic systems, physics-based modeling, hybrid approach, data scarce environments, machine learning, LightGBM, encoder-decoder, attention.

I. INTRODUCTION

GREEN technologies are crucial for Europe's goal to be carbon neutral by 2050. This objective has led to a significant increase in the installation of renewable energy systems by electric energy providers. Concurrently, this is paralleled by a rise in the transition of buildings from passive consumers to active prosumers and the emergence of net-zero energy buildings. However, a study in Japan suggests that high penetration of behind-the-meter Photo Voltaic (PV) systems might paradoxically lead to increased carbon emissions [1], mainly due to grid

imbalances caused by inaccurate forecasts. Effective forecasting enhances the management of renewable energy intermittency, improves grid reliability, optimizes energy use, and benefits investors in the energy market [2].

PV power output forecasting methods are broadly classified into data-driven (statistical), physics-based, and hybrid techniques [3]. Data-driven models fall into two categories: single-variant and multi-variant methods. Single-variant methods, such as Auto Regressive Integrated Moving Average (ARIMA) and its derivatives [4], [5], [6], use historical sequence data to identify autocorrelation and trends. Additionally, single-variant ML methods, particularly Recurrent Neural Networks (RNNs) like Long Short-Term Memory (LSTM) models, are effective in recognizing patterns in sequence data [7], [8].

Single-variant methods can sometimes miss critical exogenous factors, like weather conditions. To overcome this, more sophisticated methodologies have been developed to examine the intricate relationships between the target variable and external parameters. In the sphere of multi-variant data-driven models, diverse approaches are utilized. These encompass Artificial Neural Networks (ANN) [9], a range of regression techniques including linear and Bayesian methods [10], Support Vector Regression (SVR) models [11], tree-based methods such as Gradient Boosted Regression Trees (GBRT) [12], and advanced deep learning methodologies [13], [14]. Overall, these multi-variant models are found to be more effective than single-variant models in capturing dependencies within data.

Physics-Based Models (PBMs) utilize mathematical modeling grounded in fundamental system physics to replicate systems' behavior. These models are employed to estimate solar power output, factoring in aspects such as solar irradiance, ambient temperature, panel technology, and installation characteristics, as detailed in [15]. PBMs are especially useful in environments with limited data, where data-driven methods might be less effective. However, PBMs come with inherent challenges, including the necessity for in-depth domain knowledge and complexities in further improvement. They frequently need modifications to accommodate factors like panel degradation or environmental shifts [16]. Additionally, PBMs are typically optimized for standard testing conditions and specific panel technologies, which can limit their broader applicability. Nevertheless, recent advancements have led to the development of more versatile mathematical models capable of functioning

Received 2 December 2023; revised 30 June 2024; accepted 23 October 2024. Date of publication 28 October 2024; date of current version 24 March 2025. This work was supported by the Estonian Research Council under Grant PRG675 and Grant PRG1463. Paper no. TSTE-01341-2023. (*Corresponding author: Hossein Nourollahi Hokmabad.*)

Hossein Nourollahi Hokmabad and Oleksandr Husev are with the Electrical Power Engineering and Mechatronics, Tallinn University of Technology (TalTech), 19086 Tallinn, Harjuma, Estonia (e-mail: hossein.nourollahi@taltech.ee; oleksandr.husev@taltech.ee).

Juri Belikov is with the Software Science, Tallinn University of Technology (TalTech), 19086 Tallinn, Harjuma, Estonia (e-mail: juri.belikov@taltech.ee).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TSTE.2024.3486907>.

Digital Object Identifier 10.1109/TSTE.2024.3486907

1949-3029 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

effectively under diverse temperature and irradiance conditions, as demonstrated in [17].

Hybrid models, which combine various methodologies, have demonstrated promising results, outperforming the first two groups of models [7], [18], [19], [20]. The primary motivation for this integration is to capitalize on the strengths of different approaches while mitigating their individual limitations. The composition of these hybrid models is quite diverse. For instance, some models amalgamate multiple ML techniques to form robust systems [21], while others integrate physics-based insights with data-driven tools [22]. Furthermore, hybrid approaches sometimes incorporate signal or image processing techniques, such as wavelet analysis and sky image processing, alongside Artificial Intelligence (AI) technologies [23], [24], enhancing their analytical capabilities.

Recent advancements in solar output power forecasting have incorporated encoder-decoder models, traditionally utilized in machine translation and natural language processing, into their structural framework. A notable example is a day-ahead forecasting method employing an autoencoder-LSTM hybrid network, as discussed in [25]. RNN-based encoder-decoder models are especially valued for their ability to capture long-range dependencies within data. The inclusion of attention mechanisms further strengthens these models by allowing them to focus on specific parts of the input sequence, thereby improving both the accuracy and relevance of the forecasts. An example of such innovation is the development of an attention-based Temporal Convolutional Neural Network (TCNN) stacked model for spatio-temporal forecasting, presented in [26]. This model leverages the attention mechanism to better interpret spatial and temporal data, illustrating the flexibility and efficiency of encoder-decoder models in various forecasting scenarios.

PV panel output variability, due to weather fluctuations, challenges forecasting models' accuracy. To tackle this issue, several studies have proposed the use of classifiers to segment data based on distinct weather characteristics, thereby enabling the development of class-specific models. For example, [3] presented a hybrid classifier-regressor approach for day-ahead forecasting. In a similar vein, [27] devised a method that leverages weather variables to identify days with similar characteristics. Additionally, [28] employed SVRs to develop several specialized models catering to various weather conditions and daily time intervals. Despite advancements in solar power forecasting, several gaps still require further investigation. A significant gap exists in the study of scenarios lacking historical records. In such cases, the absence of pre-existing data at the onset can substantially impede the initial performance of ML models, highlighting the necessity for alternative methods. Moreover, although training models on weather categorization and specific weather classes typically enhances a system's overall accuracy, data-driven classification approaches are prone to classification errors and overfitting. Additionally, their reliance on initially trained weather conditions constrains their adaptability and scalability, thereby diminishing their effectiveness in diverse forecasting scenarios in different regions.

In this study, we introduce a novel day-ahead forecasting mechanism developed for behind-the-meter PV systems. The

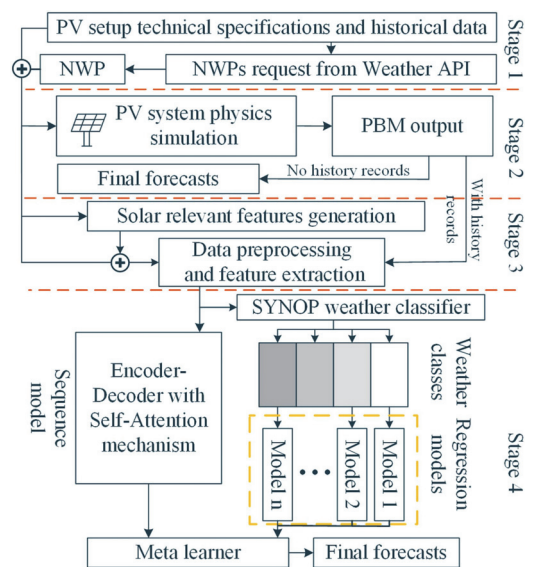


Fig. 1. Proposed framework for day-ahead solar generation forecasting.

main contributions of this research are as follows: (1) The fusion of system physics-based insights with ML techniques to address scenarios with abundant data and to alleviate cold start issues common in ML applications. This feature allows for immediate solar power predictions right after PV installation. (2) The development of an innovative ML framework that combines the strengths of regression-based models, specifically LightGBM, and sequence analysis techniques, notably a Self-Attention based Encoder-Decoder (SAED) model, to enhance forecasting accuracy for day-ahead predictions. (3) The utilization of surface synoptic weather codes (SYNOP) for weather data classification, streamlining the data clustering process, minimizing computational overhead, and reducing potential errors or overfitting risks, while also ensuring model applicability in various climates and locations. (4) Leveraging publicly accessible Numerical Weather Predictions (NWP) services to achieve consistent performance across diverse geographical locations, and to enhance the global applicability of the proposed method. (5) Investigating how the volume of collected data affects the performance of data-driven and physics-based models during system operation.

II. PROPOSED METHOD

A visual flowchart of the proposed framework is depicted in Fig. 1. Briefly, the technique can be classified into several distinct stages.

Stage 1 (data retrieval): Obtaining archived weather forecasts using the open-source NWP services, considering the specific locale of the PV setup and historical data timeline, if it exists.

Stage 2 (physics-based modeling): A PBM utilizes the installed system's technical specifications and climate data to

simulate the expected output power. If the system under investigation lacks historical data, the outputs from the PBM are considered the final forecast. Conversely, when historical data is available, the focus shifts to the development of the ML models. In this scenario, physics-based knowledge is employed to enhance the performance of the ML model.

Stage 3 (feature generation and data pre-processing): Exogenous features, like solar positioning, which are instrumental in improving the performance of ML models, have been generated. Following the integration of these features into the dataset, it undergoes a pre-processing phase. This step is essential for indigenous feature extraction.

Stage 4 (classification, regression, and sequence learning): The data instances are categorized based on their SYNOP weather codes. For each weather class, distinct regressors are employed to train custom regression models. Concurrently, selected features in the dataset are reshaped to create an appropriate data structure for training a custom SAED network.

Stage 5 (meta learner): The predictions from the regression models and the SAED model are combined using a meta-learner. This combination serves to enhance the overall accuracy of the final forecasting model and reduce the residuals from the previous stages.

A. Physics-Based Model

PBMs use fundamental principles to explain system behaviors, relying on understanding underlying physical processes. The prevalent technique for solar cell simulation is the 5-parameters model, which characterizes the solar cell behavior with a single diode equivalent circuit. Additional components inherent to this model include an irradiance controlled current source, shunt, and series resistors, given as [17]:

$$\begin{aligned} I &= I_{ph} - I_d - I_{pc} \\ &= I_{ph} - I_0 \left[\exp \left(\frac{q(V' + R_{sc}I)}{\eta\kappa T_c} \right) - 1 \right] - \frac{V' + R_{sc}I}{R_{pc}}, \end{aligned} \quad (1)$$

where I is the solar cell output current, V' is the output voltage, I_{ph} is the generated current by solar irradiance, I_d is the diode current, I_{pc} is the current of the parallel resistor (R_{pc}), R_{sc} is the series output resistor, I_0 is diode reverse saturation current, η is the ideality factor, q is the charge of the electron (1.602×10^{-19} C), κ is the Boltzmann's constant (1.381×10^{-23} J/K), and T_c is solar cell temperature in Kelvin and can be calculated as [29]:

$$T_c = T_a + \frac{T_{NOC} - 20}{800} \times SI, \quad (2)$$

where T_a is the ambient temperature (K), T_{NOC} is panel's temperature in °C in normal operation condition (extracted from panel datasheet), and SI refers to the solar irradiance (w.m^{-2}). In PV panels, solar cells are connected in series/parallel formats.

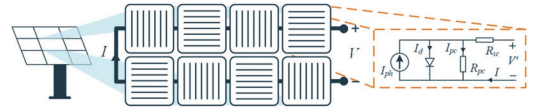


Fig. 2. PV module layout and equivalent circuit model of solar cells connected in the series format.

TABLE I
SIMULATED PV SYSTEM PARAMETERS

Parameter	Value/Name
PV model	AU Optronics PM072MW0 345W
Cell type	Monocrystalline
Inverter	Yaskawa Solectria Solar PVI 50 kW
Panels count	88
Installation type	ground based (Fix)
Azimuth angle	165°
Tilt angle	30°
Temperature model	Open-rack-glass-glass

In series style, PV module's I/V relation can be calculated as:

$$I = I_{ph} - I_0 \left[\exp \left(\frac{q(V + N_s R_{sc} I)}{N_s \eta \kappa T_c} \right) - 1 \right] - \frac{V + N_s R_{sc} I}{N_s R_{ps}}, \quad (3)$$

where $V = V' N_s$, and N_s denotes number of the cells in series in each single module. Fig. 2 shows the PV module layout connected in the series format.

The output power of a PV system is not only influenced by the characteristics of the solar panels but also by the performance of the ancillary infrastructures, like solar inverters. Consequently, the power generated by a PV system at the maximum power point (MPP) for an individual panel can be approximated as [28]:

$$P_{pv} = \left[P_{pv}^{STC} \times \frac{SI}{1000} \right] \times [1 - \gamma \times (T_j - 25)], \quad (4)$$

and,

$$T_j = T_{amb} + \frac{SI}{800} \times (N_{OCT} - 20), \quad (5)$$

where P_{pv} represents the panel's output power at MPP and P_{pv}^{STC} represents the output power in the standard test condition, γ is a temperature coefficient at MPP, T_j is the solar panel temperature (°C), T_{amb} is ambient temperature (°C), and N_{OCT} is a constant value.

To model the infrastructure of the target PV system and ensure compatibility with a wide variety of commercial solar inverters, we recommend utilizing the open-source package pvlib, developed by the U.S. Sandia National Laboratories [30]. This tool encompasses various PV cell technologies and provides the capability to design PV cells using custom PBMs. Detailed parameters of the PV cell model, along with other relevant system-specific data, are documented in Table I. For synthesizing the target system's output power, various weather parameters such as ambient temperature, global horizontal irradiance, direct normal irradiance, diffuse horizontal irradiance, wind speed,

and total accumulated precipitable water are essential. These weather parameters must be sourced from renowned global weather prediction organizations, based on the geographical location of the PV installation and corresponding time span. Subsequently, the gathered weather parameters are inputted into the PBM to predict potential PV generation for the following conditions.

B. Data Description and Preprocessing

This research examines a behind-the-meter PV setup located within a 30 km radius of Tallinn, Estonia, with a capacity of 30.6 kWp. Initially, the solar setup's output power data, measured and collected at five-minute intervals throughout 2021, 2022, and the first half of 2023. NWP data was sourced from the Open-meteo API, an open-source weather service that provides hourly weather forecasts in Europe. This API delivers forecasts that include parameters such as solar irradiance and cloud coverage. The recorded power output data was then averaged and down-sampled from five-minute intervals to hourly intervals to align with the granularity of the NWP data.

In the subsequent stage of preparing the dataset, additional features were generated. The performance of ML solar forecasting models can be enhanced by incorporating parameters like solar zenith, azimuth, and elevation, which were calculated using the geolocation of the PV region. The detailed methodologies for these calculations are discussed in [31]. Moreover, the results of PBM simulations, labeled as 'pvsim', were incorporated as a feature in the dataset for ML framework development. Autocorrelation and correlogram analysis, conducted using the Auto Correlation Function (ACF) and the Partial Auto Correlation Function (PACF), have revealed significant dependencies between the PV output power and its time lags, particularly those corresponding to the previous day's hours. Notably, auto correlations are more pronounced at lags that are multiples of the seasonal frequency, specifically 24 hours. Therefore, multiples of lag 1 and 2, such as lag 24 and lag 25, emerge as robust features that could significantly enhance the performance of ML forecasting models. Additionally, lags 47, 48, and 49, along with their corresponding subsequent day lags, also demonstrate significant correlations with the target value. Therefore, lags with correlation coefficients above 0.75 were additionally included in the dataset, with this threshold value being determined through trial and error. However, it's important to note that this threshold is not a universal design factor and may not be the best level for filtering out lag values in all cases.

To more effectively capture the inherent seasonality in the target label data, we applied sinusoidal and cosinusoidal transformations to cyclical features (C_f) such as month, day, and hour instead of using their initial values. The transformations are formulated as: $C_f^{\cos} = \cos(\frac{2\pi C_f}{\max(C_f)})$, $C_f^{\sin} = \sin(\frac{2\pi C_f}{\max(C_f)})$, where C_f^{\cos} and C_f^{\sin} represent the cosinusoidal and sinusoidal transformations of the cyclical features, respectively. Finally, the data underwent scaling using the Min-Max method, constraining the values between 0 and 1. This scaling approach helps avoid masking effects by maintaining a consistent range across all features. The scaling factors, determined based on the training

TABLE II
LAND STATION SYNOP WEATHER CODES [32]

Code	Precipitation	Description
00	no	Cloud development not observable
01	no	Clouds dissolving or becoming less developed
02	no	State of sky on the whole unchanged
03	no	Clouds generally forming or developing
51	yes	Slight drizzle
53	yes	Moderate drizzle
55	yes	Heavy drizzle
61	yes	Slight rain
63	yes	Moderate rain
71	yes	Slight continuous fall of snowflakes
73	yes	Moderate continuous fall of snowflakes
75	yes	Heavy continuous fall of snowflakes

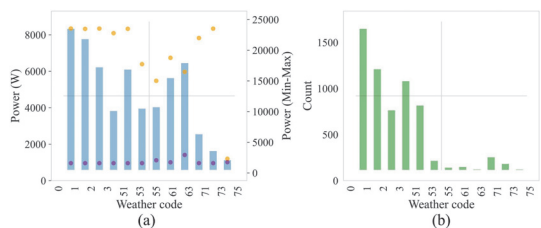


Fig. 3. (a) Average (bar plot), maximum (orange dots), and minimum (purple dots) harvested power for different weather codes in the installed PV system. (b) Number of total observed hours within the same class.

set, were then consistently applied to the test set to prevent data leakage.

C. Weather Conditions Classification

Climate variations are significant across different seasons, and many methodologies underscore the benefits of classifying weather and developing class-specific ML models. Here, we use SYNOP codes for weather categorization, a system endorsed by the world meteorological organization, applied by meteorological stations globally. This system classifies climate into 100 distinct groups. Table II lists only the weather codes observed at the investigated PV setup location. Fig. 3(a) and (b) illustrate the hourly average, minimum and maximum values of PV output power segmented by SYNOP codes and the number of observations per weather code, respectively.

This approach offers several advantages: 1) It eliminates the need for additional clustering or classification computations. 2) The codes are universally recognized and clear, ensuring consistency. 3) It is particularly suitable for scenarios where data may be limited. 4) It reduces the risk of classification overfitting and potential errors.

Furthermore, in cases where SYNOP codes are absent in the acquired NWP data, it is possible to generate these codes by utilizing weather information such as cloud coverage, temperature, precipitation, wind speed, direction, etc. Since, these weather factors are essential for accurate solar power generation forecasting, using either physics-based or data-driven models,

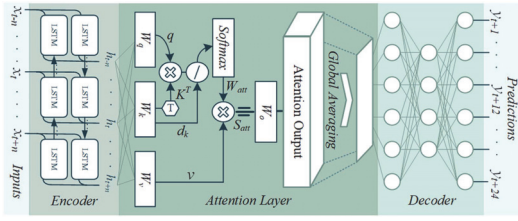


Fig. 4. Purposed SAED based solar power forecasting model.

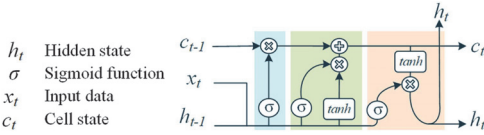


Fig. 5. LSTM units' cell structures.

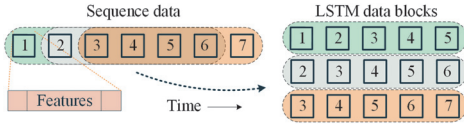


Fig. 6. Sliding window technique.

therefore, adding SYNOP codes to NWP datasets does not noticeably increase the data processing burden.

D. Sequence Forecasting Model

For sequence data analysis, a custom SAED model has been developed, which has 3 main parts: encoder, self-attention layer, and decoder, as depicted in Fig. 4.

For the encoder layer, a Bidirectional LSTM network (BiLSTM) is implemented for encoding the input sequence data. BiLSTMs are capable of processing sequence data in both forward and backward directions, thus improving the model's insight into sequential data. Fig. 5 depicts the LSTM unit cell layouts. An LSTM cell consists of four Neural Networks (NNs), each dedicated to a unique role. As seen in Fig. 5, the blue-highlighted section determines memory retention; the green section, equipped with two distinct NNs, formulates and selects long-term memory candidates; and the orange section forecasts the output using past memory and current information. In LSTM cells, the tanh and sigmoid (σ) functions have been implemented to prevent gradient issues and ensure network stability, irrespective of the memory length.

To prepare the dataset for the BiLSTM layer, transforming it into a 3-D tensor shape is essential. This transformation was achieved using a sliding window technique, as illustrated in Fig. 6. In our approach, we utilized a look-back window of 360 records, corresponding to a continuous past 15 days period. Additionally, the sliding window is designed to move 24 steps each time, providing a new data segment for each day. The

prediction horizon is set to 24 hours, aligning with the goal of day-ahead forecasting. Thus, by assuming input batch data with (I, S, F) dimensions, where I is the number of data blocks, S is the memory length, and F is the number of features. The mathematical description of the LSTM cell with \mathcal{N} hidden units is as:

$$\begin{aligned}
 f_t &= \sigma(W_f \times x_t + U_f \times h_{t-1} + b_f), & f_t \in \mathbb{R}^{\mathcal{N}}, \\
 i_t &= \sigma(W_i \times x_t + U_i \times h_{t-1} + b_i), & i_t \in \mathbb{R}^{\mathcal{N}}, \\
 o_t &= \sigma(W_o \times x_t + U_o \times h_{t-1} + b_o), & o_t \in \mathbb{R}^{\mathcal{N}}, \\
 c'_t &= \tanh(W_c \times x_t + U_c \times h_{t-1} + b_c), & c'_t \in \mathbb{R}^{\mathcal{N}}, \\
 c_t &= f_t \odot c_{t-1} + i_t \cdot c'_t, & c_t \in \mathbb{R}^{\mathcal{N}}, \\
 y_t &= h_t = o_t \odot \tanh(c_t), & h_t \in \mathbb{R}^{\mathcal{N}},
 \end{aligned} \tag{6}$$

where $W_* \in \mathbb{R}^{\mathcal{N} \times F}$ and $U_* \in \mathbb{R}^{\mathcal{N} \times \mathcal{N}}$ represent weight matrices for input data $x_t \in \mathbb{R}^F$ in timestep t and hidden state data $h_{t-1} \in \mathbb{R}^{\mathcal{N}}$ in previous timestep $t-1$, respectively, while $b_* \in \mathbb{R}^{\mathcal{N}}$ represents bias term for each gate and \odot denotes element wise product. f, i, o, c are abbreviations for *forget, input, output* gates, and *cell state*, respectively.

Before feeding the data processed by the encoder into the decoder layer, the self-attention mechanism, initially pivotal in machine language translation and computer vision [33], has been integrated between the encoder and decoder layers to enhance performance. This mechanism, inspired by human vision, has become fundamental in natural language processing and generative AI models. Its recent application in various forecasting areas, including solar output power forecasting, has yielded valuable results [34], [35], [36].

The attention mechanism allows NNs to focus on specific segments of the input sequence, akin to human selective perception. In encoder-decoder architectures, it processes the encoder's hidden states through an attention head. This head generates attention scores that pinpoint crucial information within the sequence. These scores are then utilized by the decoder network to emphasize and retain the most informative parts of the input. As a result, this selective focus significantly enhances the model's overall effectiveness. The versatility of the attention mechanism allows its adaptation across various networks dealing with sequential data.

We have implemented a self-attention mechanism, inspired by the methodologies outlined in [37], [38]. The structure of this attention mechanism is illustrated in Fig. 4. The input matrix to the self-attention layer is represented as $\mathcal{H} = \{h_{t-n}, \dots, h_t, \dots, h_{t+n}\}$, where $\mathcal{H} \in \mathbb{R}^{S \times \mathcal{N}}$. This mechanism computes three matrices from \mathcal{H} : Queries (q), Keys (k), and Values (v), using learned dense transformations. The formulas for these are $q = \mathcal{H}W_q$, $k = \mathcal{H}W_k$, and $v = \mathcal{H}W_v$, with q, k, v each belonging to $\mathbb{R}^{S \times d_{att}}$. Here, W_q, W_k , and W_v are the respective weight matrices for Queries, Keys, and Values, sized $\mathbb{R}^{\mathcal{N} \times d_{att}}$. The dimension d_{att} denotes the number of units in the dense layers. Attention weights W_{att} are derived using the dot product of Queries and transposed Keys, followed by a softmax

TABLE III
 ML MODELS HYPERPARAMETERS

Model Name	Weather Class	Name/Value
LightGBM	0	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 0.84, Lr: 0.0957, MD: 8, MCS: 26, MDIL: 14, NoE: 98, NoI: 752, NoL: 50, L1: 5.283, L2: 1.173, ES: 75}
	1	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 0.66, Lr: 0.0976, MD: 5, MCS: 87, MDIL: 14, NoE: 1198, NoI: 640, NoL: 164, L1: 4.989, L2: 0.657, ES: 20}
	2	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 1, Lr: 0.0765, MD: 10, MCS: 2, MDIL: 25, NoE: 1368, NoI: 1243, NoL: 23, L1: -, L2: -, ES: 20}
	3	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 1, Lr: 0.0764, MD: 18, MCS: 71, MDIL: 1, NoE: 367, NoI: 1400, NoL: 52, L1: -, L2: -, ES: 200}
	others	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 0.8, Lr: 0.0420, MD: 18, MCS: 10, MDIL: 2, NoE: 184, NoI: 471, NoL: 196, L1: -, L2: -, ES: 100}
SAED	-	{Number of BiLSTM layers: 1, LSTM cells: 50, Attention layer width: 100, FNN depth: 4, HL _1 size: 60, HL _2 size: 120, HL _3 size: 60, output size: 24, optimizer: Adam, Loss: MSE, Lr: 5e-4, epoch: 100, batch size: 32}
Meta learner	-	{loss: huber, penalty: elasticsearch, Lr: invscaling, alpha: 0.0016, eta0: 0.3, NoI: 50, power_t: 0.48, tol: 6e-4}
Recursive-LSTM*	-	{Number of LSTM layers: 2, LSTM cells: 60, drop out: 0.25, Dense: 1, optimizer: Adamax, Lr: 3e-4, epoch: 60, batch size: 16}
XGBoost*	-	{MD: 6, eta: 0.007, SS: 0.9, CST: 0.9, CL: 0.2, MSW: 94.81, L1: 1.63, L2: 29.7, gamma: 0.9}
LSTM*	-	{Number of LSTM layers: 2, LSTM cells: 200, drop out: 0.15, Dense: 24, optimizer: Adam, Lr: 5e-4, epoch: 200, batch size: 8}
BiLSTM*	-	{Number of BiLSTM layers: 1, LSTM cells: 50, output Dense: 24, optimizer: Adam, Lr: 3e-4, epoch: 100, batch size: 32}
LightGBM*	-	{Bt: goss, Obj: MAPE, metric: RMSE, Ff: 1, Lr: 0.0501, MD: 17, MCS: 24, MDIL: 1, NoE: 782, NoI: 900, NoL: 35, L1: -, L2: -, ES: 200}
Encoder-Decoder*	-	{ Number of BiLSTM layers: 1, LSTM cells: 50, FNN depth: 4, HL _1 size: 60, HL _2 size: 120, HL _3 size: 60, output size: 24, optimizer: Adam, Loss: MSE, Lr: 5e-4, epoch: 100, batch size: 32}

Benchmark models

Hyperparameters abbreviations: Boosting type (Bt), objective (Obj), Feature fraction (Ff), Learning rate (Lr), Max Depth (MD), Minimum Child Samples (MCS), Min Data in Leaf (MDIL), Number of Estimators (NoE), Number of Iterations (NoI), Number of Leaves (NoL), Regression Alfa (L1), Regression Lambda (L2), Early Stopping (ES), Sub Sample (SS), Minimum Child Weight (MSW), Column Sample by Tree (CST), Column Sample by Level (CL), Hidden Layer (HL).

operation as:

$$W_{att} = \text{softmax} \left(\frac{qk^T}{\sqrt{d_{att}}} \right), \quad W_{att} \in \mathbb{R}^{S \times S}, \quad (7)$$

where the squared form d_{att} is used as scaling factor. This scaling is crucial for stabilizing the gradients during the learning process. Without scaling, the dot product could grow large in magnitude, pushing the softmax function into regions where it has extremely small gradients, which is known as the softmax bottleneck. The final attention scores is obtained by matrix multiplication of v and W_{att} as:

$$S_{att} = W_{att}v, \quad S_{att} \in \mathbb{R}^{S \times d_{att}}. \quad (8)$$

The output of the attention layer (O_{att}) is then obtained by applying another learned dense transformation ($W_o \in \mathbb{R}^{d_{att} \times S}$, $b_o \in \mathbb{R}^S$) to the attention output:

$$O_{att} = S_{att}W_o + b_o, \quad O_{att} \in \mathbb{R}^{S \times d_{att}}. \quad (9)$$

Finally, O_{att} is transformed using 1-D global averaging. The global averaging is implemented to squeeze the dimension of the attention scores to be compatible with decoder layer.

In the decoder segment of the model, a Feedforward Neural Network (FNN) is utilized to process and synthesize the insights gathered from the preceding layers. This FNN comprises three hidden layers. The output layer is distinctively designed

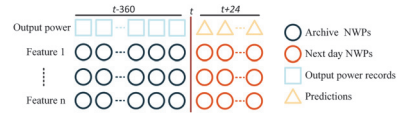


Fig. 7. SAED network input data structure.

for one-shot forecasting of the next day, featuring 24 output neurons, each representing an hour of the ensuing day. One-shot forecasting is generally preferred over recursive methods, as it is less prone to error propagation. The intricate technical details and specific characteristics of the decoder layers are comprehensively documented in Table III.

During the training phase of the SAED model, the initial 80% of the reshaped data is utilized for training, with the remaining 20% dedicated to making predictions. The model leverages five key features: *Solar Radiation Parameters*, *Cloud Coverage*, *Temperature*, *pvsim*, and *Output Power* records. Fig. 7 displays the internal structure of the input data blocks fed into the encoder layer. The training objective is to enable the model to predict PV output power by analyzing past records alongside the forthcoming 24 hours' NWP's. However, since the last 24 hours of the *Output Power* data column are the target values to be predicted, missing values in these data chunks are filled with zeros.

E. Regression Models

In our study, for each weather class a specific regression model has been developed. However, due to a limited number of occurrences in certain weather code classes in our dataset, namely '53', '55', '61', '63', '71', '73', '75', these classes were merged into the class code '51' and collectively labeled as 'others'. Consequently, this led to the formation of 5 distinct weather classes, each equipped with its own regression engine. To ensure that these models can effectively generalize across different seasons, the train datasets were shuffled.

In this study, the selected regression model is LightGBM. LightGBM incorporates two key techniques: Gradient-Based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). EFB bundles mutually exclusive features, like one-hot-encoded ones, reducing data dimensionality and speeding up model training. Conversely, GOSS optimizes instance selection for calculating information gain. Instead of using all instances, GOSS focuses on high gradient ones and randomly samples the rest, lowering computational costs. Furthermore, LightGBM discretizes continuous features into integers to efficiently construct decision trees, creating feature histograms to identify potential split points. The instances have been split in LightGBM based on the estimated variance $v_j(s)$ over subsets described as [39]:

$$v_j(s) = \frac{1}{n} \left(\frac{v_l(s)}{n_l^j(s)} + \frac{v_h(s)}{n_h^j(s)} \right), \quad (10)$$

$$v_l(s) = \left(\sum_{x_i \in A_l} g_i + \frac{1}{\alpha - \beta} \sum_{x_i \in B_l} g_i \right)^2, \quad (11)$$

$$v_h(s) = \left(\sum_{x_i \in A_h} g_i + \frac{1}{\alpha - \beta} \sum_{x_i \in B_h} g_i \right)^2, \quad (12)$$

where $x_i \in \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^m$, m is the number of features in the feature space χ^m , and n is the total number of samples in the subsets $(A \cup B)$. A refers to the subset of $\alpha\%$ of all data samples with highest gradient and B represents the subset of the $\beta\%$ randomly selected instances from the remaining data samples. ψ represents the potential point of splitting based on feature j . subsets are divided based on ψ as: $A_l = \{x_i \in A : x_{ij} \leq \psi\}$, $B_l = \{x_i \in B : x_{ij} \leq \psi\}$, $A_h = \{x_i \in A : x_{ij} > \psi\}$, $B_h = \{x_i \in B : x_{ij} > \psi\}$. Furthermore, n_l^j refers to number of instances in the subset $A_l \cup B_l$ and n_h^j refers to number of instances in the subset $A_h \cup B_h$. g_i is the negative gradient of the instance x_i , and term $\frac{1}{\alpha - \beta}$ is used to normalize the sum of gradients. The training procedure for both branches of the proposed model is summarized in Algorithm 1.

F. Meta Learner

In our framework, we have selected a linear regression model trained using a Stochastic Gradient Descent Regressor (SGDR) [40] as the meta learner. This type of regression is adept at fitting by minimizing a regularized empirical loss through Stochastic Gradient Descent (SGD). The choice of SGDR is due to its online learning capability, which enables iterative

Algorithm 1: Proposed ML Branches' Training Procedure.

Input: dataset $\mathcal{D}^{train} = \{(x_i, y_i)\}$, hyperparameters $\mathfrak{H}_{\mathfrak{M}}$, $\mathfrak{M} \in \{\text{LightGBM}, \text{SAED}, \text{SGDR}\}$, batch size B
Output: weights $\tilde{w}_{\mathfrak{M}}$, $\mathfrak{M} \in \{\text{SAED}, \text{SGDR}\}$ and $\mathcal{F}_{\mathcal{J}}(\cdot)$, $\mathcal{F} \in \{\text{LightGBM}\}$, $\mathcal{J} \in \{ '00', '01', '02', '03', 'others' \}$

- 1: Identify \mathcal{J} for all x_i : $\mathcal{J}_i \leftarrow \text{SYNOPE}(x_i)$
- 2: Generate class-based datasets:
 $\mathcal{D}_{\mathcal{J}}^{train} \leftarrow \text{Classifier}(\mathcal{D}^{train}, \mathcal{J}_i)$
- 3: **for all** \mathcal{J} **do**
- 4: $\mathcal{F}_{\mathcal{J}} \leftarrow \mathcal{F}(\mathcal{D}_{\mathcal{J}}^{train}, \mathfrak{H}_{\mathcal{F}})$
- 5: **end for**
- 6: Initialize \tilde{w}_{SAED} with random weights:
- 7: $\{\tilde{w}_{\overrightarrow{\text{BiLSTM}}}, \tilde{w}_{\overleftarrow{\text{BiLSTM}}}, \tilde{w}_{\text{att}}, \tilde{w}_{\text{fnn}}, \tilde{b}_{\text{fnn}}\} \sim \mathcal{R}$
- 8: **for** $\text{epoch} = 1$ **to** Epochs **do**
- 9: Initialize total loss: $\mathcal{L}_T = 0$
- 10: **for all** batches $\mathcal{B} \subseteq \mathcal{D}^{train}$ **do**
- 11: **for all** pairs (x_i, y_i) **in** \mathcal{B} **do**
- 12: $\overrightarrow{h}_i, \overleftarrow{h}_i \leftarrow \text{BiLSTM}(\overrightarrow{\tilde{w}_{\text{BiLSTM}}}, \overleftarrow{\tilde{w}_{\text{BiLSTM}}}, x_i)$
- 13: $W_{\text{att}}^i \leftarrow \text{Attention}(\overrightarrow{h}_i, \overleftarrow{h}_i)$
- 14: $S_{\text{att}}^i \leftarrow W_{\text{att}}^i \nu_i$
- 15: $O_{\text{att}}^i \leftarrow S_{\text{att}}^i W_o^i + b_o^i$
- 16: $\hat{y}_i \leftarrow \sigma(W_{\text{fnn}} O_{\text{att}}^i + b_{\text{fnn}})$
- 17: Accumulate batch loss: $\mathcal{L}_T \leftarrow \mathcal{L}_T + \mathcal{L}(\hat{y}_i, y_i)$
- 18: **end for**
- 19: Average batch loss: $\mathcal{L}_B \leftarrow \frac{\mathcal{L}_T}{B}$
- 20: Calculate gradients:
 $\mathcal{G} \leftarrow \text{Backpropagation}(\mathfrak{M}, \mathcal{L}_B)$
- 21: $\tilde{w}_* \leftarrow \text{UpdateWeights}(\mathfrak{M}, \mathcal{G}, \text{LearningRate})$
- 22: **end For**
- 23: **end For**

model updates with new data inputs. This adaptability is crucial for accommodating changes resulting from environmental variations or alterations in system components. While there are other viable methods, such as weighted averaging, the use of SGD for training strikes offers a balance between performance and computational efficiency, making it a preferable choice for dynamic and evolving forecasting scenarios.

G. Hyper Parameters Optimization

The hyperparameter optimization stage for tree-based models like LightGBM in this study was conducted using Bayesian optimization, as cited in [41]. This technique is particularly effective because it leverages information from previous iterations to guide the search, making it more efficient than Grid or Random search methods, especially for models with a large number of hyperparameters. Decision trees are susceptible to overfitting, especially with smaller datasets, which makes careful hyperparameter selection imperative and thus the Bayesian optimization shows to outcome with better results than Random and Grid search methods. For all models, the hyperparameters have been optimized after conducting 100 iterations with 10 initial points. In contrast, for training the SAED, the SGDR meta learner, and RNN-based benchmark models, the Grid search algorithm has

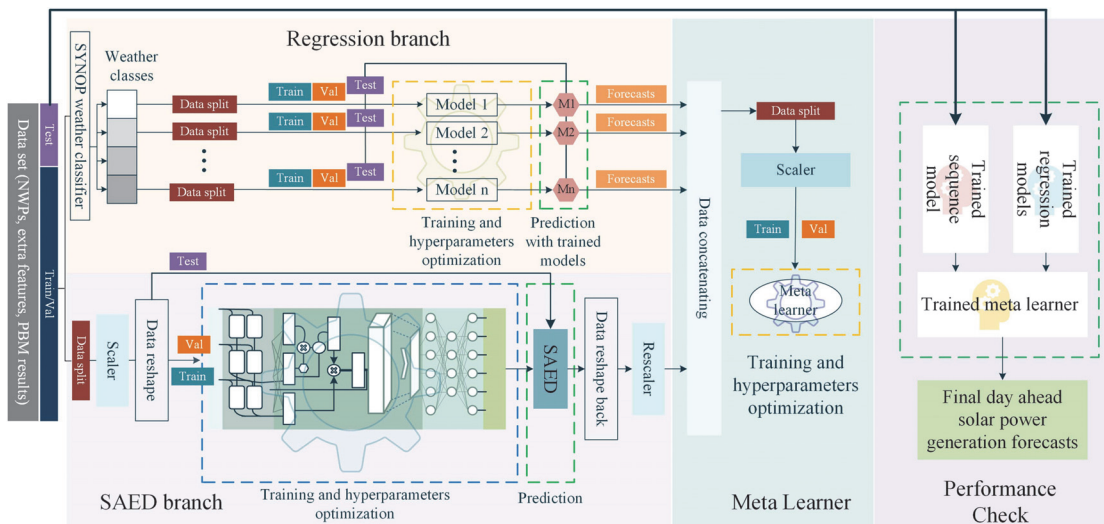


Fig. 8. The detailed pipeline of the proposed hybrid ML model.

been employed. This choice was made because the number and range of hyperparameters in these models are limited, making Grid search a suitable choice.

H. Hybrid ML Models Pipeline

Fig. 8 summarizes the detailed pipeline related to the data segmentation, scaling, training, validation, and test evaluation of the proposed hybrid ML model. At the initial stage, the dataset is divided into training and testing sets. The test dataset is exclusively reserved for the final hybrid ML model performance evaluation. During the training of the regression models, the main training dataset is segmented based on weather codes. Once data instances are classified, the subsets are further divided into train-validation and test sub-datasets. Importantly, as we plan to train a sequence model later, the data is split sequentially without shuffling. The subset designated as the test sub-dataset is used to generate outputs for training a meta learner from learned regression models represented as M1, M2, ..., Mn in Fig. 8. At this stage, the train-validation datasets are shuffled and split into sub-train and sub-validation datasets. Specific regression models for each class are then trained and undergo hyperparameter optimization. After the training and optimization of the models, each regression model processes its respective sub-test dataset to produce initial predictions. Following these procedures for all classes, we compile a dataset that is ready to be integrated with the outputs of the sequence model through a meta learner.

In the second branch, the main dataset undergoes a similar division procedure as the regression branch. After this division, a Min-Max scaler is applied to normalize the sub-train data, and the same scaler is then used to transform the sub-validation dataset. The test set is similarly scaled using the same fitted scaler to maintain consistency. Following data scaling, the datasets are separately prepared and reshaped to facilitate the training,

hyperparameter optimization, and testing phases. Subsequently, the trained SAED model processes the sub-test dataset to generate initial predictions. These predictions are then rescaled and integrated with the results from the regression models. During the meta learner training stage, the outputs from previous stages are divided into training and validation subsets. The meta learner is then trained, and its hyperparameters are simultaneously optimized.

In Fig. 8, the 'Performance Check' block illustrates the principle of model operation after training all models. At this stage, the pre-trained models are fed with test data, which was set aside at the beginning. It is important to note that the test data undergoes the same classification/scaling procedures as in the regression stage and the same scaling/data reshaping as in the SAED branch. These steps are not repeated in the flowchart to avoid over-complexity. Ultimately, the outputs from both branches are combined by a meta learner, resulting in 24 final forecasts that are generated and delivered.

III. NUMERIC RESULTS

In this section, we separately discuss the performance of the PBM, LightGBM, and the SAED model. Following this, we compare the overall performance of these models with selected benchmarks. The development of these models was facilitated using the Keras 2.10 and Scikit-learn 1.3.10 libraries, running on Python 3.10.7. All computational processes were carried out on a desktop system equipped with an Intel Core i7-7700K CPU (4.20 GHz) and 32 GB of RAM.

For the training of regression models and for metrics reports for all models, only data instances during daytime were used. Furthermore, forecasts for the subsequent day for all models are generated at midnight after acquiring NWP data. The test set spans 8760 hours (365 days), covering the latter half of 2022

and the first half of 2023. This period allows for the analysis of the models' ability to capture seasonality and adaptation to both extremely low and high solar generation, due to the unique daylight conditions in northern European countries. In late spring and summer, day lengths can extend up to 21 hours, significantly longer than in regions closer to the equator. Conversely, during autumn and winter, day lengths can shorten to as low as 5 hours, making these seasons less suitable for evaluating forecast performance. This comprehensive year-round analysis helps in understanding the models' performance under varying solar conditions. Table III indicates the best hyperparameters sets for all the models.

A. Evaluation Metrics

Regression and forecasting model evaluations commonly utilize metrics like normalized Root Mean Square Error (nRMSE), normalized Mean Absolute Error (nMAE), coefficient of determination (R^2) score, Symmetric Mean Absolute Percentage Error (sMAPE), and Mean Absolute Percentage Error (MAPE). Notably, MAPE and sMAPE can pose challenges, especially when both the actual target and the forecasted value are zero, potentially leading to undefined values. To mitigate this issue, MAPE and sMAPE in this study are calculated using a method recommended in [3]. The formulas for these metrics are detailed accordingly.

$$\text{nRMSE} = \frac{1}{y_{\max}} \sqrt{\frac{1}{n} \sum_{h=1}^n (y_h - y'_h)^2}, \quad (13)$$

$$\text{nMAE} = \frac{1}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{y_{\max}}, \quad (14)$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{TSS}, \quad (15)$$

$$\text{MAPE} = \frac{100}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{\bar{y}}, \quad (16)$$

$$\text{sMAPE} = \frac{100}{n} \sum_{h=1}^n \frac{|y_h - y'_h|}{(\bar{y} + \bar{y}')/2}, \quad (17)$$

where n is number of forecasting data samples, y and y' are actual and forecasted values, respectively. y_{\max} is the maximum amount of the target value in the corresponding test set. Also, $SS_{\text{res}} = \sum_{h=1}^n (y_h - y'_h)^2$ and $TSS = \sum_{h=1}^n (y_h - \bar{y})^2$ are sum of squares residuals and total sum of squares, respectively. \bar{y} and \bar{y}' stands for the means of actual PV output power values, and the mean of forecasted PV output power values, respectively.

B. Benchmark Models

The performance of our proposed model was assessed in comparison to a spectrum of foundational models, including LSTM, recursive LSTM, BiLSTM, XGBoost, LightGBM, a basic encoder-decoder, averaging, and the Persistence model. Furthermore, we have compared the performance of our

TABLE IV
FORECAST PERFORMANCE OF LIGHTGBM MODELS

Weather class	nMAE	MAPE %	sMAPE %	nRMSE	R^2 score
0	0.029	9.22	10.13	0.052	0.967
1	0.048	15.68	16.95	0.075	0.931
2	0.082	26.88	28.62	0.120	0.819
3	0.059	35.61	34.92	0.098	0.764
others	0.077	33.61	35.07	0.115	0.729

TABLE V
COMPARISON OF FORECAST PERFORMANCES BETWEEN SUNRISE AND SUNSET PERIODS

Model	nMAE	MAPE %	sMAPE %	nRMSE	R^2 score
Proposed	0.043	22.62	21.62	0.089	0.898
Average	0.046	22.82	22.16	0.091	0.885
SAED	0.048	24.71	23.23	0.096	0.876
Classified LightGBM	0.050	25.54	24.63	0.101	0.864
LightGBM-LSTM*	0.050	25.18	24.99	0.098	0.860
Recursive-LSTM*	0.053	26.41	26.20	0.110	0.855
XGBoost*	0.051	26.36	26.36	0.108	0.855
LightGBM*	0.053	26.64	26.47	0.109	0.853
BiLSTM*	0.056	27.88	26.93	0.114	0.829
Encoder-Decoder*	0.053	26.77	26.41	0.111	0.853
LSTM*	0.057	29.18	28.08	0.116	0.844
PBM	0.070	36.03	34.08	0.134	0.725
Persistence	0.084	43.13	43.13	0.174	0.584

proposed model with state-of-the-art solutions documented in the literature. To ensure equitable benchmarking, all these models underwent the same optimization process as our proposed model. Notably, during the training of the benchmark models, the 'pvsim' column, which represents physics-based insights, was included to isolate its impact on the comparative performance of the ML models. The Persistence model employs a simple forecasting method, predicting that the PV power output for any given hour will replicate the output from the same hour on the previous day. Additionally, the average model generates forecasts by averaging the outputs from the two branches proposed in our model.

C. Physics-Based Model Performance Evaluation

The performance of PBM has been reported and compared with other data-driven methods in Table V. Although these mathematical models might not rival the performance of data-driven methods in every aspect, they hold significant value for initializing new systems. Another advantage of incorporating a PBM in forecasting tools is its ability to effectively handle weather conditions not previously seen in historical datasets. In contrast, models relying solely on ML techniques may struggle when encountering unfamiliar data instances in the test set. Fig. 9 presents the daily sum of simulated output energy derived from the PBM and the daily sum of actual recorded values, along with their residuals, during the last 4.5 months of the test period.

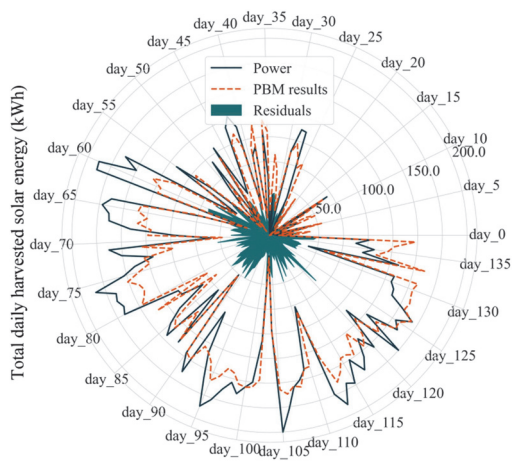


Fig. 9. Daily summed real output energy values and daily summed physics-based simulations outcomes and their residuals.

D. Regression Models Performance Evaluation

Table IV outlines the performance of LightGBM models across weather classifications. The highest accuracy is in class ‘0’, representative of sunny conditions, with a MAPE score of 9.22% and 96.7% proficiency in capturing variable changes. However, performance drops notably in weather classes ‘2’, ‘3’, and ‘others’, with class 3’s MAPE reaching 35.61% due to the unpredictability of cloud formations, which is not fully addressed by NWP.

Despite these limitations, LightGBM models perform satisfactorily overall. Note that while MAPE for class ‘3’ is the highest, other metrics indicate its superior performance over class ‘others’. This highlights MAPE’s tendency to favor higher forecasts, leading to potential underestimation and suggests the need for using multiple metrics in model evaluation. Thus, the regressor for class ‘3’, despite a tendency to overestimate, generally produces more accurate predictions than the ‘others’ class.

E. SAED Performance Evaluation

The core feature of the encoder-decoder structure applied in this study is the self-attention mechanism. The effectiveness of an attention mechanism depends on its ability to identify key segments of the sequence data, thereby guiding the decoder to focus on these areas, which enhances model performance. To assess this, evaluating the attention weights is instrumental in determining whether meaningful attention is being formed within the sequence data and if the model is learning to concentrate on significant parts of the training data.

Fig. 10 illustrates the attention weights for the training data during the model’s training process. Initially, there is no significant focus on particular features. However, as training progresses through more epochs, the attention mechanism learns to distinguish specific parts of the data by increasing attention weights and scores on them. These scores are pivotal for the decoder’s

performance and allow for the use of a relatively simple FNN in the decoder part while still achieving considerable results. The performance of the proposed SAED model was compared with a simplified version of the same encoder-decoder model (Encoder-Decoder*), which lacked the self-attention mechanism. The outcomes confirm that the inclusion of the attention mechanism improves various metrics, thus system general performance. The detailed metrics and their improvements are enumerated in Table V.

F. Proposed Model Performance Analysis

The performance metrics of the proposed model are detailed in Table V. The proposed model demonstrates commendable day-ahead forecasting performance during daytime hours, with an nMAE of 0.052, nRMSE of 0.043, MAPE of 22.62%, sMAPE of 21.62% and an R^2 score of 0.898. Notably, the performance of the ‘Average’ meta learner closely approximates that of the chosen meta learner, suggesting that the results from the two parallel branches are sufficiently accurate to employ a simple method like averaging as a combination head.

Additionally, the SAED model shows superior performance compared to the BiLSTM*, simple Encoder-Decoder*, and Recursive-LSTM* models, underscoring the effectiveness of incorporating the attention mechanism. Furthermore, the performance of the XGBoost* model, with a nMAE of 0.051 and sMAPE of 26.36%, is comparable to that of the applied LightGBM regressors. This suggests that the XGBoost regressor could also be a viable option within the framework [42].

The performance of the model across 15 consecutive days is presented in Fig. 11. It is apparent that the model’s performance varies, underperforming on certain days, which highlights areas for further improvement. However, the overall efficacy for day-ahead predictions is satisfactory. During purely sunny days, the model’s forecasting is near ideal, and it also performs well on mostly sunny days with minimal cloud coverage. Challenges arise in accurately predicting output during days with highly variable weather conditions. This issue is compounded by a decrease in the accuracy of NWP under dynamic weather conditions, affecting the model’s forecasting ability. Moreover, on rainy and cloudy days, where power generation is considerably lower and more variable, the model still provides relatively close forecasts of the target values, which should be highlighted as well.

Fig. 12 offers a comparative analysis of the actual total solar energy generation versus the forecasted day-ahead energy production and the prediction residuals. Compared to Fig. 9, there is a significant reduction in residuals, underscoring the importance of integrating ML tools with PBMs. Accurate predictions of cumulative available energy for the next day are crucial for behind-the-meter energy management systems, as their optimization algorithms rely more on aggregate renewable energy availability than on hourly predictions [43].

The ability of models to capture seasonal trends and manage seasonality parameters is essential for precise forecasting. Reliable and robust models should consistently maintain their performance within an acceptable range throughout the year and across different seasons. To validate the effectiveness of our model in

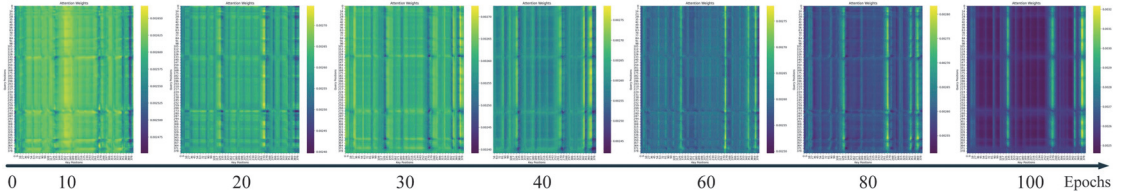


Fig. 10. Attention formation by increasing training epochs.

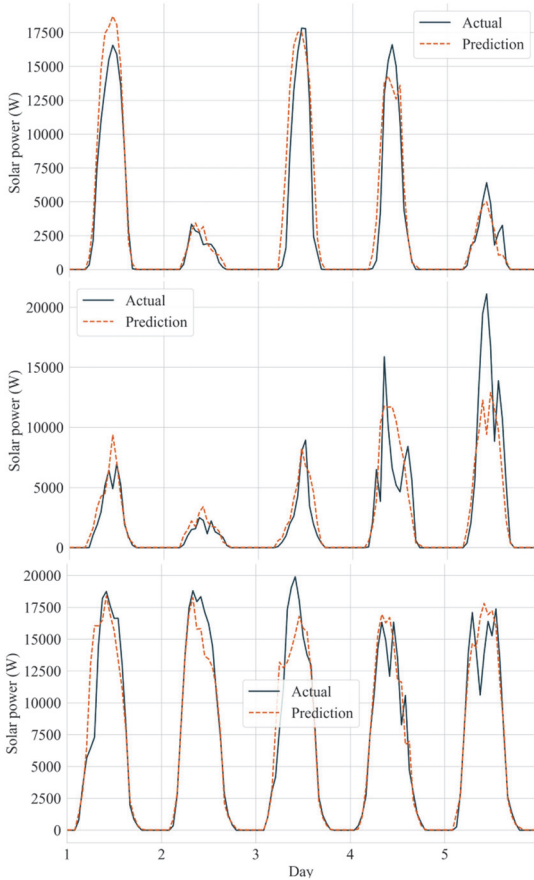


Fig. 11. Proposed forecasting framework comparison with measured data.

forecasting solar power generation across varying seasons, we conducted a detailed seasonal comparison of model accuracy. Various metrics were calculated separately for each month to assess the model’s accuracy in different seasonal contexts, providing a clearer insight into its strengths and weaknesses under various weather conditions and day lengths. These metrics were then averaged for each season and compared against benchmark models. The seasonal performance data are compiled in Tables VI to IX.

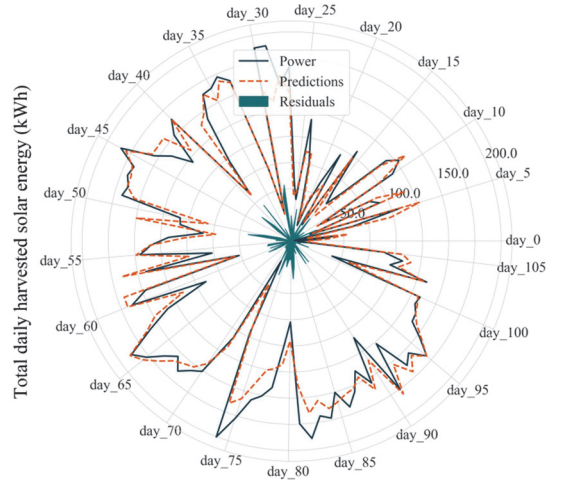


Fig. 12. Daily summed real output energy values and daily summed proposed model predictions and their residuals.

TABLE VI
COMPARISON OF MODELS’ PERFORMANCES DURING SPRING (MARCH, APRIL, MAY)

Model	mMAE	MAPE %	sMAPE %	nRMSE	R ² score
Proposed	0.041	22.28	21.78	0.084	0.908
Average	0.043	22.48	21.92	0.086	0.890
SAED	0.045	24.61	23.39	0.091	0.883
Classified LightGBM	0.047	25.69	25.40	0.097	0.871
LightGBM-LSTM*	0.049	26.11	25.95	0.096	0.860
Recursive-LSTM*	0.049	25.74	25.88	0.099	0.864
LightGBM*	0.051	26.93	27.06	0.102	0.855
BiLSTM*	0.052	27.04	26.53	0.105	0.854
Encoder-Decoder*	0.048	26.35	26.38	0.100	0.859
LSTM*	0.053	28.92	28.04	0.106	0.849
XGBoost*	0.049	26.16	26.67	0.102	0.858
PBM	0.058	31.46	32.42	0.112	0.825
Persistence	0.080	44.94	45.07	0.174	0.560

Our proposed model, which incorporates three distinct techniques for forecasting solar power generation, consistently outperforms benchmark models in most cases and across various seasons. However, like all models, it struggles significantly during cold and snowy days to maintain accuracy. This decreased

TABLE VII
COMPARISON OF MODELS' PERFORMANCES DURING SUMMER (JUNE, JULY, AUGUST)

Model	nMAE	MAPE %	sMAPE %	nRMSE	R ² score
Proposed	0.049	22.95	21.46	0.096	0.888
Average	0.049	23.15	22.39	0.097	0.879
SAED	0.051	24.81	23.06	0.100	0.868
Classified LightGBM	0.052	25.39	23.85	0.104	0.857
LightGBM-LSTM*	0.051	24.25	24.02	0.099	0.860
Recursive-LSTM*	0.057	27.07	26.51	0.120	0.846
LightGBM*	0.055	26.35	25.88	0.115	0.850
BiLSTM*	0.060	28.72	27.32	0.124	0.841
Encoder-Decoder*	0.058	27.19	26.43	0.121	0.846
LSTM*	0.061	29.44	28.11	0.126	0.839
XGBoost*	0.053	26.61	26.05	0.113	0.852
PBM	0.081	40.60	35.73	0.155	0.624
Persistence	0.087	41.32	41.19	0.173	0.608

TABLE VIII
COMPARISON OF MODELS' PERFORMANCES DURING AUTUMN (SEPTEMBER, OCTOBER, NOVEMBER)

Model	nMAE	MAPE %	sMAPE %	nRMSE	R ² score
Proposed	0.034	37.25	36.35	0.076	0.843
Average	0.036	37.67	37.59	0.079	0.836
SAED	0.036	38.72	39.23	0.082	0.823
Classified LightGBM	0.038	42.26	40.85	0.087	0.797
LightGBM-LSTM*	0.037	41.55	41.27	0.086	0.817
Recursive-LSTM*	0.038	42.06	41.81	0.088	0.784
LightGBM*	0.039	43.07	43.24	0.092	0.762
BiLSTM*	0.037	43.02	40.86	0.087	0.798
Encoder-Decoder*	0.040	47.17	46.28	0.093	0.737
LSTM*	0.039	49.23	47.53	0.091	0.752
XGBoost*	0.039	46.09	47.28	0.090	0.767
PBM	0.041	56.85	52.48	0.099	0.653
Persistence	0.066	85.50	80.73	0.162	0.311

TABLE IX
COMPARISON OF MODELS' PERFORMANCES DURING WINTER (DECEMBER, JANUARY, FEBRUARY)

Model	nMAE	MAPE %	sMAPE %	nRMSE	R ² score
Proposed	0.013	47.26	48.49	0.050	0.787
Average	0.014	49.13	51.62	0.051	0.752
SAED	0.014	52.65	53.22	0.052	0.734
Classified LightGBM	0.017	58.76	60.90	0.059	0.629
LightGBM-LSTM*	0.015	54.21	55.18	0.054	0.717
Recursive-LSTM*	0.015	55.18	56.05	0.055	0.690
LightGBM*	0.017	61.90	62.34	0.060	0.615
BiLSTM*	0.016	56.06	57.63	0.056	0.656
Encoder-Decoder*	0.016	55.82	56.87	0.055	0.679
LSTM*	0.017	60.39	58.31	0.058	0.635
XGBoost*	0.017	59.81	61.53	0.059	0.620
PBM	0.029	113.45	88.32	0.090	0.179
Persistence	0.029	104.96	108.42	0.108	0.181

TABLE X
COMPARISON OF MODELS' PERFORMANCES WITH STATE OF THE ART

Model	nMAE	MAPE %	sMAPE %	nRMSE	R ² score
Proposed	0.043	22.62	21.62	0.089	0.898
[3]	0.047	25.13	24.57	0.097	0.872
[44]	0.045	22.32	21.73	0.088	0.883
[45]	0.055	28.01	27.29	0.114	0.846

accuracy is primarily due to snow coverage on panels and the very low energy generation during such conditions, which deviate from typical weather patterns and expected power outputs. Addressing the inclusion of these phenomena in forecasting models is beyond the scope of this work. Despite the challenges posed by these extreme conditions, our model still outperforms the benchmark models, confirming its superiority and reliability even in less-than-ideal scenarios. Furthermore, all models perform best during the spring and summer months, when weather conditions are relatively stable, and the accuracy of NWP is higher compared to winter and autumn. This seasonal variation highlights the impact of environmental consistency on forecasting accuracy, as stable conditions simplify the predictive modeling process and reduce the likelihood of anomalies that could disrupt model performance.

Additionally, the potential of the proposed technology is compared with cutting-edge technologies, and the outcomes are summarized in Table X. Overall, our proposed solution outperforms its competitors in many aspects. However, it is important to note that [44] achieves slightly better results in terms of nMAE and sMAPE. Despite this, our model, which integrates diverse methodologies and combines elements of data-driven models and PBM, delivers outcomes that are reliable, accurate, and robust compared to state-of-the-art approaches. This demonstrates the efficacy of blending different analytical techniques to enhance forecasting performance. Nevertheless, there are still areas that require further investigation.

IV. CONCLUSION

Accurate solar power forecasting is pivotal for efficient energy management, grid stability, and the pursuit of carbon neutrality. Addressing this need, our study introduces a hybrid day-ahead solar power forecasting model for behind-the-meter applications, blending PBMs and ML techniques. The approach is versatile, catering to setups with extensive historical data and new installations without past records. This model is designed to be versatile, accommodating both installations with extensive historical data and those without past records, addressing cold start issues effectively.

Our findings suggest that decisions regarding the reliance on PBMs or ML models as the definitive forecasting solution should not be made until at least 12 to 15 months after initiating data collection and model training. This timeline allows for adequate adaptation to the model's learning algorithms, though it may vary based on regional differences and varying weather conditions. For instance, areas with less variability in weather

parameters might benefit from the ML model's outputs sooner than the 12th month. Such variability highlights the importance of considering local contextual and environmental factors when deploying ML predictions in diverse settings.

In the proposed method, adaptability to different climates is ensured by using open-access NWP and SYNOP weather codes for streamlined classification. The framework comprises two models: a regression approach using LightGBM and self-attention based encoder decoder branch, enhanced by a SGDR as a meta-learner. In a test period, the model achieved a MAPE of 22.62%, sMAPE of 21.62%, nMAE of 0.043, nRMSE of 0.089, and R^2 score of 0.90, thereby outperforming benchmark and state-of-the-art models. Despite its efficacy, the model showed limitations under variable weather conditions, indicating a need for improved accuracy in weather predictions, particularly for solar radiation and cloud coverage.

REFERENCES

- [1] A. Okuyama, S. Yoo, J. Kumagai, A. R. Keeley, and S. Managi, "Questioning the Sun: Unexpected emissions implications from residential solar photovoltaic systems," *Resour. Conservation Recycling*, vol. 176, 2022, Art. no. 105924.
- [2] M. Khorasany, A. S. Gazafroudi, R. Razzaghi, T. Morstyn, and M. Shafie-khah, "A framework for participation of prosumers in peer-to-peer energy trading and flexibility markets," *Appl. Energy*, vol. 314, 2022, Art. no. 118907.
- [3] M. Nejati and N. Amjadi, "A new solar power prediction method based on feature clustering and hybrid-classification-regression forecasting," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1188–1198, Apr. 2022.
- [4] A. Shadab, S. Ahmad, and S. Said, "Spatial forecasting of solar radiation using ARIMA model," *Remote Sens. Appl.: Soc. Environ.*, vol. 20, 2020, Art. no. 100427.
- [5] V. Kushwaha and N. M. Pindoriya, "A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast," *Renewable Energy*, vol. 140, pp. 124–139, 2019.
- [6] Y. Li, Y. Su, and L. Shu, "An ARMAX model for forecasting the power output of a grid connected photovoltaic system," *Renewable Energy*, vol. 66, pp. 78–89, 2014.
- [7] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs," *IEEE Trans. Ind. Inform.*, vol. 17, no. 3, pp. 1873–1881, Mar. 2021.
- [8] Q. Li, Y. Xu, B. S. H. Chew, H. Ding, and G. Zhao, "An integrated missing-data tolerant model for probabilistic PV power generation forecasting," *IEEE Trans. Power Syst.*, vol. 37, no. 6, pp. 4447–4459, Nov. 2022.
- [9] X. Meng, F. Gao, T. Xu, K. Zhou, W. Li, and Q. Wu, "Inverter-data-driven second-level power forecasting for photovoltaic power plant," *IEEE Trans. Ind. Electron.*, vol. 68, no. 8, pp. 7034–7044, Aug. 2021.
- [10] M. Bozorg, A. Bracale, P. Caramia, G. Carpinelli, M. Carpita, and P. De Falco, "Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting," *Protection Control Modern Power Syst.*, vol. 5, no. 1, pp. 1–12, 2020.
- [11] U. K. Das et al., "Optimized support vector regression-based model for solar power generation forecasting on the basis of online weather reports," *IEEE Access*, vol. 10, pp. 15594–15604, 2022.
- [12] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Sol. Energy*, vol. 150, pp. 423–436, 2017.
- [13] A. Mellit, A. M. Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renewable Energy*, vol. 172, pp. 276–288, 2021.
- [14] S. A. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short-and long-term solar irradiance forecasting for Islamabad," *Renewable Energy*, vol. 198, pp. 51–60, 2022.
- [15] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Appl. Energy*, vol. 283, 2021, Art. no. 116239.
- [16] C. Baldus-Jeursen et al., "Snow losses for photovoltaic systems: Validating the marion and townsend models," *IEEE J. Photovolt.*, vol. 13, no. 4, pp. 610–620, Jul. 2023.
- [17] T. Ma, W. Gu, L. Shen, and M. Li, "An improved and comprehensive mathematical model for solar photovoltaic modules under real operating conditions," *Sol. Energy*, vol. 184, pp. 292–304, 2019.
- [18] N. Rahimi et al., "A comprehensive review on ensemble solar power forecasting algorithms," *J. Elect. Eng. Technol.*, vol. 18, no. 2, pp. 719–733, 2023.
- [19] A. Agga, A. Abbou, M. Labbadi, Y. El Houm, and I. H. O. Ali, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Syst. Res.*, vol. 208, 2022, Art. no. 107908.
- [20] J. Wang, Y. Zhou, and Z. Li, "Hour-ahead photovoltaic generation forecasting method based on machine learning and multi objective optimization algorithm," *Appl. Energy*, vol. 312, 2022, Art. no. 118725.
- [21] J. Wang, H. Zhong, X. Lai, Q. Xia, Y. Wang, and C. Kang, "Exploring key weather factors from analytical modeling toward improved solar power forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1417–1427, Mar. 2019.
- [22] S. E. Haupt et al., "Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting," *Energies*, vol. 13, no. 8, 2020, Art. no. 1979.
- [23] Z. Wang, L. Wang, C. Huang, and X. Luo, "A hybrid ensemble learning model for short-term solar irradiance forecasting using historical observations and sky images," *IEEE Trans. Ind. Appl.*, vol. 59, no. 2, pp. 2041–2049, Mar./Apr. 2022.
- [24] M. Q. Raza, N. Mithulananthan, J. Li, K. Y. Lee, and H. B. Gooi, "An ensemble framework for day-ahead forecast of PV output power in smart grids," *IEEE Trans. Ind. Inform.*, vol. 15, no. 8, pp. 4624–4634, Aug. 2019.
- [25] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model," *IEEE Trans. Ind. Appl.*, vol. 56, no. 6, pp. 7185–7192, Nov./Dec. 2020.
- [26] H. Kim and D. Lee, "Probabilistic solar power forecasting based on bivariate conditional solar irradiance distributions," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 2031–2041, Oct. 2021.
- [27] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2253–2262, Sep. 2015.
- [28] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, Jul. 2014.
- [29] H. Tian, F. Mancilla-David, K. Ellis, E. Muljadi, and P. Jenkins, "A cell-to-module-to-array detailed model for photovoltaic panels," *Sol. Energy*, vol. 86, no. 9, pp. 2695–2706, 2012.
- [30] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: A python package for modeling solar energy systems," *J. Open Source Softw.*, vol. 3, no. 29, 2018, Art. no. 884.
- [31] I. Reda and A. Andreas, "Solar position algorithm for solar radiation applications," *Sol. Energy*, vol. 76, no. 5, pp. 577–589, 2004.
- [32] Environment and Climate Change Canada, *MANOBS - Manual of Surface Weather Observation Standards*, 8th edition, amendment 2, Gatineau, QC: Environ. Climate Change Canada, 2023.
- [33] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, 2015.
- [34] S. Sharda, M. Singh, and K. Sharma, "RSAM: Robust self-attention based multi-horizon model for solar irradiance forecasting," *IEEE Trans. Sustain. Energy*, vol. 12, no. 2, pp. 1394–1405, Apr. 2021.
- [35] T. Hu, H. Ma, H. Liu, H. Sun, and K. Liu, "Self-attention-based machine theory of mind for electric vehicle charging demand forecast," *IEEE Trans. Ind. Inform.*, vol. 18, no. 11, pp. 8191–8202, Nov. 2022.
- [36] J. Hong, J. Park, and S. Park, "StackDA: A stacked dual attention neural network for multivariate time-series forecasting," *IEEE Access*, vol. 9, pp. 145955–145967, 2021.
- [37] J. Liang and W. Tang, "Ultra-short-term spatiotemporal forecasting of renewable resources: An attention temporal convolutional network-based approach," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3798–3812, Sep. 2022.
- [38] J. Qiu, B. Wang, and C. Zhou, "Forecasting stock prices with long-short term memory neural network based on attention mechanism," *PLoS One*, vol. 15, no. 1, 2020, Art. no. e0227222.
- [39] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 1–12.

- [40] T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, p. 116.
- [41] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, vol. 25, pp. 1–9.
- [42] H. N. Hokmabad, O. Husev, D. Vinnikov, J. Belikov, and E. Petlenkov, "Day-ahead PV output power forecasting utilizing boosting recursive LightGBM-LSTM framework," in *Proc. 2023 IEEE PES Innov. Smart Grid Technol. Europe*, 2023, pp. 1–5.
- [43] H. N. Hokmabad, O. Husev, J. Kurnitski, and J. Belikov, "Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications," *Sustain. Energy, Grids Netw.*, vol. 38, 2024, Art. no. 101385.
- [44] C. Yu, J. Qiao, C. Chen, C. Yu, and X. Mi, "TFEformer: A new temporal frequency ensemble transformer for day-ahead photovoltaic power prediction," *J. Cleaner Prod.*, vol. 448, 2024, Art. no. 141690.
- [45] M. Aslam, S.-J. Lee, S.-H. Khang, and S. Hong, "Two-stage attention over LSTM with Bayesian optimization for day-ahead solar power forecasting," *IEEE Access*, vol. 9, pp. 107387–107398, 2021.



Hossein Nourollahi Hokmabad (Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering and integrated circuit design from University of Tabriz, Iran, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with the Tallinn University of Technology (TalTech), Estonia. His research focuses on the application of machine learning, deep learning, and data-driven methods in next-generation energy management systems, smart grids, and edge-AI.



Oleksandr Husev (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in industrial electronics from Chernihiv State Technological University, Chernihiv, Ukraine, in 2007 and 2008, respectively. He defended Ph.D. thesis in the Institute of Electrodynamics of the National Academy of Science of Ukraine, Kyiv, Ukraine, in 2012. He is Senior Researcher and projects leader with the Department of Electrical Power Engineering and Mechatronics, Tallinn University of Technology (TalTech), Tallinn, Estonia and is an Associate Professor with home

University (Chernihiv National Polytechnic University, Chernihiv, Ukraine). He has authored or coauthored over 200 publications and is the holder of several patents including European patents. He is recognized as a Best Young Scientist in both Taltech and home Universities. He was the recipient of the Estonian National Research Award in Technical Sciences in 2024. His research interests include Power Electronics systems, Design of novel topologies, control systems based on a wide range of algorithms, including modeling, design, and simulation. Applied design of power converters and control systems and application.



Juri Belikov (Senior Member, IEEE) received the B.Sc. degree in mathematics from Tallinn University, Tallinn, Estonia, in 2006, and the M.Sc. and Ph.D. degrees in computer and systems engineering from Tallinn University of Technology, Tallinn, Estonia, in 2008 and 2012, respectively. From 2015 to 2017, he was a Postdoctoral Fellow in the Faculty of Mechanical Engineering and Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion—Israel Institute of Technology, Haifa, Israel. He is currently an Tenured Associate Professor with the Department

of Software Science, Tallinn University of Technology. His research interests include with the edge of nonlinear control theory, IT, and energy domain.

Publication III

N. Shabbir, K. Vassiljeva, H. Nourollahi Hokmabad, O. Husev, E. Petlenkov, and J. Belikov, 'Comparative Analysis of Machine Learning Techniques for Non-Intrusive Load Monitoring', *Electronics*, vol. 13, no. 8, 2024.

Article

Comparative Analysis of Machine Learning Techniques for Non-Intrusive Load Monitoring

Noman Shabbir ^{1,2}, Kristina Vassiljeva ³, Hossein Nourollahi Hokmabad ^{2,4}, Oleksandr Husev ^{2,*}, Eduard Petlenkov ³ and Juri Belikov ⁴

- ¹ FinEst Centre for Smart Cities, Tallinn University of Technology, 19086 Tallinn, Estonia; noman.shabbir@taltech.ee
 - ² Department of Electrical Power Engineering & Mechatronics, Tallinn University of Technology, 19086 Tallinn, Estonia; hossein.nourollahi@taltech.ee
 - ³ Department of Computer Systems, Tallinn University of Technology, 11712 Tallinn, Estonia; kristina.vassiljeva@taltech.ee (K.V.); eduard.petlenkov@taltech.ee (E.P)
 - ⁴ Department of Software Science, Tallinn University of Technology, 11712 Tallinn, Estonia; juri.belikov@taltech.ee
- * Correspondence: oleksandr.husev@taltech.ee

Abstract: Non-intrusive load monitoring (NILM) has emerged as a pivotal technology in energy management applications by enabling precise monitoring of individual appliance energy consumption without the requirements of intrusive sensors or smart meters. In this technique, the load disaggregation for the individual device is accrued by the recognition of their current signals by employing machine learning (ML) methods. This research paper conducts a comprehensive comparative analysis of various ML techniques applied to NILM, aiming to identify the most effective methodologies for accurate load disaggregation. The study employs a diverse dataset comprising high-resolution electricity consumption data collected from an Estonian household. The ML algorithms, including deep neural networks based on long short-term memory networks (LSTM), extreme gradient boost (XgBoost), logistic regression (LR), and dynamic time warping with K-nearest neighbor (DTW-KNN) are implemented and evaluated for their performance in load disaggregation. Key evaluation metrics such as accuracy, precision, recall, and F1 score are utilized to assess the effectiveness of each technique in capturing the nuanced energy consumption patterns of diverse appliances. Results indicate that the XgBoost-based model demonstrates superior performance in accurately identifying and disaggregating individual loads from aggregated energy consumption data. Insights derived from this research contribute to the optimization of NILM techniques for real-world applications, facilitating enhanced energy efficiency and informed decision-making in smart grid environments.

Keywords: non-intrusive load monitoring; load disaggregation; pattern recognition; machine learning; deep learning



Citation: Shabbir, N.; Vassiljeva, K.; Nourollahi Hokmabad, H.; Husev, O.; Petlenkov, E.; Belikov, J. Comparative Analysis of Machine Learning Techniques for Non-Intrusive Load Monitoring. *Electronics* **2024**, *13*, 1420. <https://doi.org/10.3390/electronics13081420>

Academic Editors: Di Wu, Jing Li, Xianmin Wang and Mingliang Zhou

Received: 15 March 2024

Revised: 4 April 2024

Accepted: 7 April 2024

Published: 9 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The rising demand for an increased proportion of renewable energy resources (RES) in the coming decades, driven by the cost-effectiveness and environmental benefits of cleaner energy production, is expected to follow an upward trajectory [1]. Although RES contributes positively to sustainability and environmental concerns, their intermittent nature poses challenges in the residential energy sector [2,3]. Therefore, it is crucial to strike a balance between demand and supply to effectively manage these energy resources. The inclusion of shiftable and non-essential loads in the residential sector, such as electric vehicles (EV) and battery energy storage systems (BESS), can play a pivotal role in optimizing energy management and enhancing system flexibility. The strategy involves scheduling these loads to coincide with the availability of RES-like photovoltaic (PV) energy. This approach not only reduces energy consumption costs and promotes sustainability and

self-reliance but also augments the penetration of renewable energy [4]. Referred to as energy flexibility (EF), this adaptability is essential for transitioning towards eco-friendly and efficient energy grids. A noteworthy development in this context is the emergence of demand-side energy aggregators, which contribute to balancing demand and supply by minimizing peak loads during periods of high demand, thereby ensuring stability in power systems, and facilitating EF [5].

The EF is generally referred to as the customer's capacity to adjust or modify behavior based on energy demand, production variations, weather conditions, and user or grid requirements [6,7]. Several devices in the household are known as shiftable loads such as EVs, washing machines, dishwashers, etc. These devices are not essential and could be used at a later time, therefore, referred to as shiftable/movable devices. Another prevalent definition focuses on the earliest start time and ending time of shiftable devices. Traditional EF characterization involved installing smart meters on residential devices and continuously monitoring data, which, although straightforward, could be costly and slow it also raised concerns about data privacy [8].

A new approach, non-intrusive load monitoring (NILM), has been proposed as an alternative. NILM observes the usage patterns of devices based on their current signals, eliminating the necessity for smart meters [9]. The total energy consumption of the user is given to the NILM model as an input and then the device usage times are extracted, this method is known as load disaggregation. This makes NILM an essential tool for demand-side management (DSM) and EF applications. Although the most precise method to measure device usage is through energy meters on individual devices, this approach is not the most practical [10]. The integration of data-driven technologies, such as machine learning (ML), into NILM has enhanced its efficiency. A detailed review of the NILM method is given in [11]. NILM solutions can be categorized into supervised and unsupervised learning [12]. In supervised learning, the model is trained using a dataset, followed by testing and verification. On the other hand, unsupervised learning involves the model extracting information from data and forming clusters without prior training sets [13]. While unsupervised learning is faster and more convenient, it lacks the accuracy of supervised learning [14]. Several ML-based methods, including K-nearest neighbor (kNN), neural networks, support vector machine (SVM), deep learning (DL), and event matching classification, have been proposed in supervised NILM. There are many studies that have also incorporated statistical methods such as particle swarm and Markov chain models. In Table 1, a comparison of previous studies with this study has been presented.

Table 1. Comparison of the current study with existing literature.

Study	Year	Place	Method Used for NILM	Dataset	Avg. Efficiency (%)
[15]	2019	China	Particle swarm	1 year	94.2
[16]	2023	Indonesia	Random Forests	1 year	99
[17]	2020	India	Markov Chain	31 days	94
[5]	2021	Estonia	Extreme Gradient Boost (XgBoost)	3 years	97.2
[18]	2022	Malaysia	K-NN, SVM, Ensemble	30 days	98.8
[19]	2020	Iran	SVM	1 week	98.2
[20]	2021	Indonesia	Convolutional Neural Networks (CNN)	1 month	98
[21]	2023	Italy	Random Forests	27 months	96.3
[22]	2024	Spain	Long Short-Term Memory Networks (LSTM)	7 months	98
[23]	2023	Greece	Recurrent Neural network (RNN)	10 days	97
[24]	2023	Canada	LSTM	2 days	98
[25]	2023	UAE	Current Waveform Features with Rule-Based Set Theory (CRuST)	1 month	96
This Study	2024	Estonia	DTW-KNN, Logistic Regression (LR), XgBoost, RNN-LSTM	1 year	98

The NILM method has been used for anomaly detection at the appliance level by incorporating machine learning [26]. In another study [8], NILM is utilized for the event

matching of devices. This study was based on the Pecan Street dataset, and it used a deep learning algorithm for this event matching. In [27], the NILM technique is used to identify the load patterns, and then later, these patterns are used to improve the accuracy of the load forecasting. A NILM-based solution has been proposed for energy management in microgrids [28]. Furthermore, the solution also provides input for the electricity market based on the load characterization by NILM. The results indicated that using this technique the energy costs and load curtailment can be reduced. In [29], a NILM-based algorithm has been proposed for the monitoring of loads in the power distribution network. This technique consists of a neural network and improves accuracy by 5%.

In most of the literature presented above, there are several methods used in NILM modeling. However, the accuracy of these NILM models is challenging as there are many device variations, different manufacturers with different power ratings and device operating modes. The inclusion of ML and DL methods improves this performance significantly but still, these methods require larger datasets of reference device signals which is problematic. Therefore, there is a gap in the research studies about the comparison of different ML and DL algorithms on the accuracy of the NILM technique. Moreover, the impact of the size of the dataset on the performance of NILM is also of interest. This paper tries to fill this gap by evaluating the performance of several ML and DL algorithms employed in NILM. These models are designed based on a real-life dataset measured in an Estonian household for the whole year. These are the main contributions of this work:

- Thorough comparative analysis of ML Techniques for NILM, revealing optimal methodologies for load disaggregation.
- Utilization of diverse dataset from an Estonian household for comprehensive evaluation of ML algorithms.
- Implementation and Evaluation of LSTM, XgBoost, LR, and DTW-KNN models, highlighting XgBoost's superior performance.
- Insightful evaluation metrics application includes accuracy, precision, recall, and F1 score for nuanced assessment.
- Identification of XgBoost as the most effective model for load disaggregation, offering practical implications for enhancing energy efficiency.

The rest of the article is structured as follows: Section 2 provides detailed background information about NILM and the ML and DL methods used in this research. The case study of the Estonian household and the development of these NILM models are presented in Section 3. The results and discussion are given in Section 4. Finally, the conclusion and future works are summarized in Section 5.

2. Non-Intrusive Load Monitoring (NILM)

Non-Intrusive Load Monitoring (NILM) is a progressive approach for estimating individual appliance operating states and their energy consumption based on household total electrical load measured at a single point. It involves acquiring and disaggregating the overall electricity usage, offering a simple and cost-effective means of monitoring appliances' operation and energy consumption formulated as:

$$P_{total}(t) = \sum_{k=1}^K p_k(t) + e(t) \quad (1)$$

where

$P_{total}(t)$ is the power consumed by all appliances,

$p_k(t)$ —power consumed by the k th appliance,

$e(t)$ —error or difference between aggregate meter reading and the sum of actual power consumption.

The examination relies on the measurement of voltage and current waveforms taken at the electrical service entrance (ESE). These data serve as the basis for deducing the operational conditions and power consumption of each individual load. Load signatures, also known as load features, are derived from these waveforms, providing measurable

parameters that reveal information about the nature and operating status of individual appliances. Appliances can be categorized into distinct types based on their load signatures, shaping the approach to disaggregation:

- Type-I appliances, such as toasters and boilers, exhibit a straightforward ON/OFF state.
- Type-II devices like washing machines and ovens operate with multiple (finite) number of operating states with recognizable patterns.
- Type-III devices, presented by dimmer lights, belong to continuously variable devices (CVD), presenting a challenge in disaggregation due to their constantly varying consumption.
- Type-IV, devices that are constantly in operation and have different energy consumption modes like smoke detectors and refrigerators.

Given the diversity outlined above, developing an accurate yet broadly applicable NILM system is a challenging task. Consequently, many algorithms are designed to focus on identifying only the most significant appliances. This strategic approach acknowledges the complexity of capturing the varied operational signatures across different appliance types while aiming to provide targeted and effective load disaggregation. The goal is to strike a balance between accuracy and generalization, ensuring that the NILM system can reliably identify and monitor key appliances without becoming overly intricate and challenging to implement [30,31]. A general NILM process can be presented in four phases and observed in Figure 1.

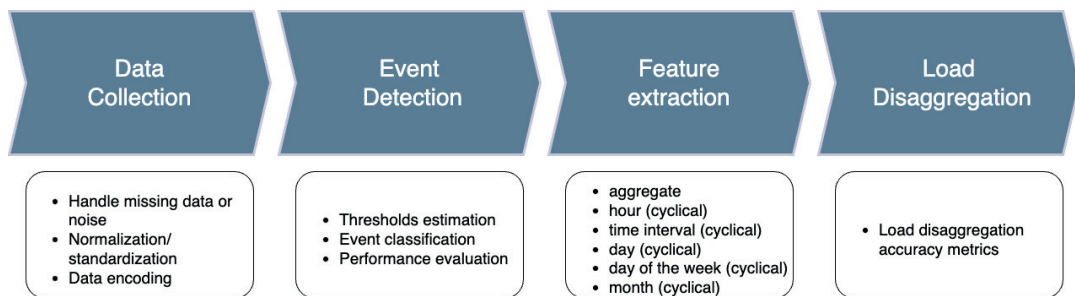


Figure 1. The general flow of the NILM technique.

2.1. Data Collection

The initial step in any NILM algorithm involves data acquisition, typically obtained from smart meters. The crucial question in load disaggregation is determining the optimal data collection frequency for smart meters to ensure accurate appliance identification and power estimation. The trade-off between high and low data frequencies significantly impacts NILM algorithm effectiveness. High-resolution measurements, often exceeding 1 Hz, can extract transient features crucial for identifying appliances with similar power consumptions, particularly during state transitions. On the other hand, excessively small data frequencies limit feature extraction to steady-state characteristics, proving it insufficient for differentiating appliances with comparable power usage.

The sampling frequency, an essential factor in data pre-processing, varies based on the appliance signature of interest, with researchers recognizing the utility of both low-frequency and high-frequency signatures. High-frequency data, however, require high-end hardware, additional data storage, and have transmission problems, and thereby increasing costs. Recent NILM solutions strategically balance algorithmic efficiency and performance across a diverse range of appliances, often favoring low-frequency signals to achieve satisfactory results [32].

The algorithms differ significantly in their approach to handling data at the collection stage. The below-mentioned algorithms have the following differences: DTW-KNN excels

in time series classification, accommodating speed variations but lacks explicit handling of missing data or noise. XgBoost robustly handles tabular data, automatically adapting to missing data and outliers, despite needing careful tuning and pre-processing. Logistic Regression, suitable for binary classification, demands meticulous pre-processing, especially for categorical data, and lacks inherent handling of missing values and noise. LSTM networks, an expert at processing sequential data, are robust to noise but may struggle with lengthy sequences, necessitating truncation or summarization and requiring numerical input. Based on the above mentioned each algorithm offers unique strengths, and optimal performance relies on meticulous data pre-processing and tuning tailored to the task and data characteristics.

2.2. Event Detection

Within the domain of NILM, event detection serves as a crucial task, focusing on detecting state transition actions generated by appliances. The event detection module identifies instances of state transitions in the aggregated power signal, characterizing actions like ON/OFF switches, changes in appliance speed and mode alternations. Challenges in event detection arise from high fluctuations, long transitions and near-simultaneity, and misidentification of events can lead to decreased accuracy and increased computational complexity in NILM methods. The event detection module employs various models, including expert heuristics, probabilistic models, and matched filter models, to identify events in the aggregate signal, with a subsequent focus on exploring different signatures for effective NILM research, including steady-state features extracted from low-frequency sampled data around the event detection window. Despite their ease of extraction, steady-state features face challenges of feature overlapping and susceptibility to power disturbances, highlighting the ongoing efforts to enhance NILM methodologies [30].

In the event detection phase, the process begins with threshold estimation, where a specific value is set or calculated dynamically to identify when an event, such as an appliance turning on or off, has occurred. This threshold is typically based on changes in power consumption and aims to minimize both false positives and negatives. Following the detection of events, they are classified into different categories, often corresponding to individual appliances. The performance of this event classification is then evaluated using various metrics such as precision, recall, and F1-score. These metrics assess the accuracy of the classification in terms of the proportion of correctly identified events, the proportion of actual events that were missed, and the balance between precision and recall, respectively.

2.3. Feature Extraction

Effective NILM methods necessitate distinctive features or signatures that capture the unique behaviors of appliances, facilitating the differentiation of various types of appliances. These features are derived from the distinctive power consumption patterns exhibited by individual appliances and are utilized to identify or recognize corresponding appliances from aggregated signals. Two main categories of features employed in NILM are transient features and steady-state features. Transient features, extracted from the transition process between two steady states, require high-frequency data acquisition by smart meters, typically exceeding 1 Hz. Event detection methods separate the transition process from overall measurements, posing a challenge to accurately capture the start and end of transitions. The steady-state features, on the other hand, encompass variables such as active power, reactive power, current, and voltage waveform, and can be extracted from conventional smart meter data without the need for high-frequency sampling. Although steady-state features are commonly used, determining the number of states remains challenging [30,33].

Feature extraction is a crucial step that involves processing the collected data to extract meaningful information. Features mentioned above (see Figure 1) can be used individually or in combination to improve the performance of the systems. For example, “aggregation” refers to the total energy consumption data collected from the main power line. It serves as the primary input for NILM systems. Date/time features can capture

daily/weekly/monthly/annual patterns in energy usage. Time intervals can refer to the duration for which an appliance is used, as different appliances tend to be used for different lengths of time [31,32]. These features can be used individually or in combination to improve the performance of NILM systems. The choice of features often depends on the specific characteristics of the problem at hand, such as the number of appliances, the sampling rate of the data, and the availability of training data [33,34].

2.4. Load Disaggregation

The final stage in the process is load disaggregation, where the identified features and patterns are used to determine the individual energy consumption and operational states of specific appliances within a building. During the load disaggregation phase, machine learning or pattern recognition algorithms, previously trained on labeled datasets in the earlier stages, are applied to the real-time or historical aggregated energy data [35]. These algorithms use the learned patterns and features to attribute portions of the total energy consumption to specific appliances. The complexity of load disaggregation lies in the fact that multiple appliances may be operating simultaneously, and their energy signatures may overlap. Advanced machine learning models are often employed to handle these challenges and improve the accuracy of disaggregation. The choice of algorithm often depends on specific characteristics of the data and the complexity of the task, including the number and types of appliances, the intricacy of energy usage patterns, and the availability of labeled training data [36].

Recognizing the diverse nature of energy consumption patterns, a strategy involving multiple algorithms has been chosen. NILM studies have explored both supervised and unsupervised approaches. Supervised methods, utilized in our approach, require a labeled dataset with sub-metered appliances. However, this kind of dataset may not always be available. On the other hand, unsupervised methods can be applied without prior knowledge of the environment. Nevertheless, users are required to validate identified appliance patterns. As our data are labeled, we primarily use supervised methods in our approach. Dynamic Time Warping (DTW) is employed for its ability to measure similarity between sequences, providing flexibility in capturing dynamic variations in energy consumption. The K-NN algorithm leverages the proximity of data points to classify patterns, contributing a robust method for identifying similarities in energy signatures. XgBoost, a powerful ensemble learning technique, excels in handling complex relationships and boosting predictive performance [37]. Lastly, LSTM is chosen for its effectiveness in discerning patterns in high-dimensional spaces [38]. By integrating these algorithms, we aim to enhance the accuracy and versatility of our load disaggregation process, addressing the complexities inherent in energy consumption data.

3. Machine Learning Techniques

ML revolutionized NILM by providing transparency and precision in energy consumption analysis. ML algorithms excel at analyzing vast datasets and uncovering hidden patterns within the energy signal. They can learn from historical data to identify unique signatures of individual appliances, even when operating simultaneously. ML's dynamic nature allows it to adapt continuously to evolving usage patterns and seasonal variations, ensuring sustained accuracy over time. Ultimately, ML transforms raw energy data into actionable insights, empowering users to optimize energy management.

3.1. Dynamic Time Warping with K-Nearest Neighbor (DTW-KNN)

DTW is a crucial algorithm for time series classification, where the objective is to train a model capable of accurately predicting the class of a time sequence within the labeled dataset [39]. The K-NN algorithm is commonly employed for this task, with a modification using the DWT metric instead of the classic Euclidian distance. DTW accommodates variations in length and speed between the compared time series, making it particularly effective for capturing patterns in energy consumption over time [40]. Despite

its efficiency, the challenge lies in the time complexity of DTW, especially for large datasets with lengthy sequences [41]. However, understanding the nuances of DTW allows for necessary adjustments to enhance the algorithm’s speed, ensuring practical and efficient time series classification in the context of NILM.

For one-dimensional time series denoted as $f(x(i))$ and $f(y(j))$, where i and j represent time points in series, and x and y are vectors, characterized by their Euclidian distances [42]. The DTW algorithm involves creating a local cost matrix D storing pairwise distances between x and y . The algorithm seeks an optimal warping path under certain constraints using dynamic programming, determining the DTW distance as the minimum accumulated distance normalized by the length of the optimal warping path. This alignment process minimizes the “distance” between the two-time series is presented in Equation (2):

$$D_{i,j} = d(x_i, y_j) + \min \begin{cases} D_{i-1,j-1} & (\text{match}) \\ D_{i-1,j} & (\text{insertion}), \\ D_{i,j-1} & (\text{deletion}) \end{cases} \tag{2}$$

where $d(x_i, y_j) = |x_i - y_j|$.

The k-nearest neighbors (K-NN) nonparametric statistical algorithm relies on k training samples in proximity to the feature space as input. The classification of an object is based on the most frequently occurring class among the identified k nearest points. The parameter k denotes the number of nearest neighbors influencing the classification process, and the selection of an appropriate k is a nuanced yet crucial step for optimizing the model’s performance [43].

The integration of DTW and K-NN in a combined approach is motivated by the distinctive strengths of each method. This integration yields a more robust and accurate predictive model, specially tailored for applications in time series analysis. Essentially, the synergy between DTW and K-NN capitalizes on DTW’s efficacy in capturing temporal nuances and K-NN’s proficiency in pattern classification based on similarity. This combined approach facilitates a more comprehensive analysis of time series data, proving particularly beneficial when dealing with complex and dynamic patterns [43].

3.2. Extreme Gradient Boosting

XgBoost is a highly efficient machine learning algorithm known for its effectiveness in predictive modeling tasks. As a gradient-boosting algorithm in the ensemble learning family, XgBoost excels in capturing intricate patterns and relationships within energy consumption data [44]. Its strength lies in accurately identifying and distinguishing between energy signatures of diverse appliances, making it invaluable in scenarios with complex and evolving consumption behaviors [44,45]. Operating as a tree ensemble model with k trees XgBoost predicts outcomes for data samples (x_i, y_i) through a defined expression [46]:

$$\hat{y}_i = F_k(x_i) = F_{k-1}(x_i) + f_k(x_i), \tag{3}$$

where

$F_{k-1}(x_i)$ is the prediction result of previous $k - 1$ trees,
 $f_k(x_i)$ — k -th decision tree.

The algorithm’s objective function involves a cost function, assessing the error between predicted and actual values.

$$F_{obj} = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{j=1}^k \Omega(f_j) \tag{4}$$

The regularization term incorporates the L1-norm, preventing overfitting by penalizing the number of leaf nodes, and the L2-norm, penalizing leaf node weights. Each iteration introduces a new tree, and the objective function is approximated using first and second-order gradients.

3.3. Logistic Regression

Logistic regression plays a pivotal role in load disaggregation within NILM systems for binary classification tasks, determining the ON/OFF states [47]. In this context, logistic regression models are trained using labeled data where the state of each appliance is known. Features extracted from the aggregated power signal, such as voltage, current, and frequency, serve as input variables for the logistic regression model. The model learns the relationship between these features and the probability of an appliance being in the on or off state [48].

During inference, the trained logistic regression model is applied to real-time aggregated power data to predict the probability of each appliance being ON or OFF. By setting a threshold probability, appliances are classified as either ON or OFF, providing valuable insights into individual appliance usage patterns. The performance of the regression-based load disaggregation model is then evaluated using metrics such as accuracy and precision and recall, with iterative optimization techniques like feature selection and hyperparameter tuning applied to enhance model efficacy.

3.4. Long Short-Term Memory Networks (LSTM)

The LSTM algorithm has become one of the essential tools in NILM due to its ability to overcome the limitations of traditional Recurrent Neural Networks (RNNs), especially in handling long-term dependencies and gradient vanishing issues [49]. LSTMs are particularly favored for NILM tasks because they excel at capturing the inherent long-term dependencies present in time series data. Equipped with forget, input, and output gates, the LSTM architecture provides precise control over information flow within each memory block, allowing for the retention of relevant information while discarding extraneous data.

The hidden layer of an LSTM network is a crucial component comprising gated units or cells, which work in tandem to generate both the cell's output and internal state (see Figure 2). Consisting of four interconnected elements, including three logistic sigmoid gates and one hyperbolic tangent (tanh) layer, LSTMs exhibit a sophisticated mechanism for controlling information flow within the cell [50]. The forget gate, employing a sigmoid activation function, determines the relevance of information from the previous cell state, aiding in the removal of obsolete data. Subsequently, the input gate combines current input with the previous hidden state, filtering pertinent information and generating new candidate values for the cell state through a tanh layer. Finally, the output gate normalizes cell state values and produces the final output, emphasizing LSTMs' capability to retain long-term dependencies and regulate information flow effectively [51].

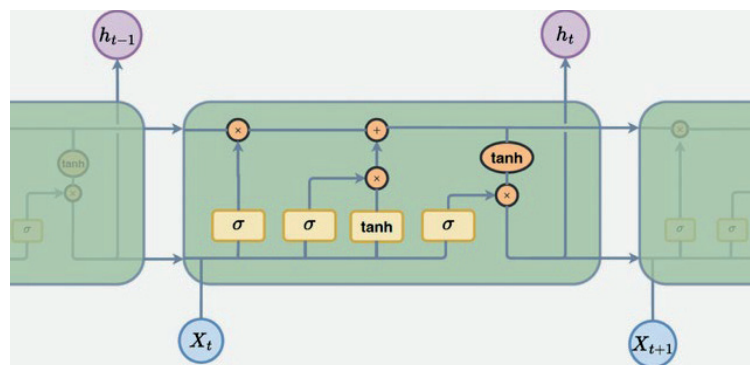


Figure 2. Structure of the LSTM Network.

It is essential not to overlook the need for additional signal processing when integrating neural networks (NNs) into applications. Circular timestamps provide a cyclic representation of time, which is beneficial for handling periodic data such as daily or

seasonal patterns. They enable neural networks, especially LSTM models, to better capture recurring patterns in tasks like time series forecasting and energy consumption modeling [52]. When using circular timestamps with LSTMs, it is crucial to encode timestamps as angles on a unit circle and design networks to handle circular sequences effectively. This approach enhances LSTM models' ability to accurately learn cyclic patterns across diverse domains, offering a compact yet powerful representation of time-related data [53].

3.5. Performance Indicators

Evaluating NILM systems requires careful consideration since a single metric cannot capture all its nuances. Although metrics like mean squared error (MSE) and false positive/negative rates offer insights into overall accuracy, the evaluation should extend to specific appliance identification metrics such as precision, recall, and F1-score [22]. These metrics provide a granular understanding of how well the system distinguishes between individual appliances, which is essential for practical implementation in real-world scenarios [10,43].

$$\text{Precision} = \frac{TP}{TP + FP'} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN'} \quad (6)$$

$$F1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

A confusion matrix is a fundamental technique in machine learning that serves as a concise summary of a classification algorithm's performance. It provides a tabular layout of the correct and incorrect predictions made by the classifier, mapping these predictions to the original classes of the data. This matrix offers crucial insights that go beyond simple accuracy metrics, which is especially valuable when dealing with imbalanced datasets or multiple classes. In the matrix, columns denote predicted values, while rows represent actual values. This arrangement offers a clear visualization of the model's accuracy and the patterns of its errors across all classes simultaneously. This structured grid aids in understanding the classifier's performance by comparing the correct and incorrect predictions for each class.

The evaluation, however, does not end at core performance metrics. Computational efficiency, adaptivity, and data handling diversity must also be considered. Metrics such as execution time and memory usage shed light on the system's computational demands, crucial for real-time applications and resource-constrained environments. Flexibility metrics gauge the system's ability to adapt to new appliances and environmental changes, ensuring its relevance and applicability over time. Finally, scalability and robustness metrics assess how well the system performs across diverse datasets and under varying conditions, offering a comprehensive picture of its reliability and generalizability.

4. Case Study of an Estonian Household

4.1. Exploratory Data Analysis

In this study, forecasting algorithms were developed using load data from a household in Estonia. The specific residence is situated in Tallinn city, comprising two levels, four rooms, and holding a "C" energy rating. With a total area of around 100 square meters, it was designed for two adults and one child. The data was measured using the Emporia Gen 2 3-PHASE device with 16 Sensors. Data collection spanned from August 2021 to August 2022, achieving an accuracy rate with an error margin below 5%. Measurements were taken at 15-min intervals. The analysis included various household appliances such as a dishwasher, vacuum cleaner, television, stereo, sauna, ventilation system, refrigerator, lighting fixtures, electric stove, and washing machine. Additionally, the house featured a heating system, water heater, and electric heating floor. The key DC power consumers in the residence comprised interior and exterior lights, multiple phone and laptop chargers, a TV and sound system, and a floor heater.

The data for the entire year is illustrated in Figure 3, showcasing separately measured AC and DC loads within the household. The combined average value loads hours around 3 kW. The peak recorded load, reaching approximately 19 kW, occurred in February during the winter months when heating demands were at their peak. In Estonia, winter spans from November to March, typically witnessing higher energy consumption. Conversely, during the summer months between May and August, energy consumption drops significantly as heating demands diminish. The monthly energy consumption throughout the year is shown in Figure 4.

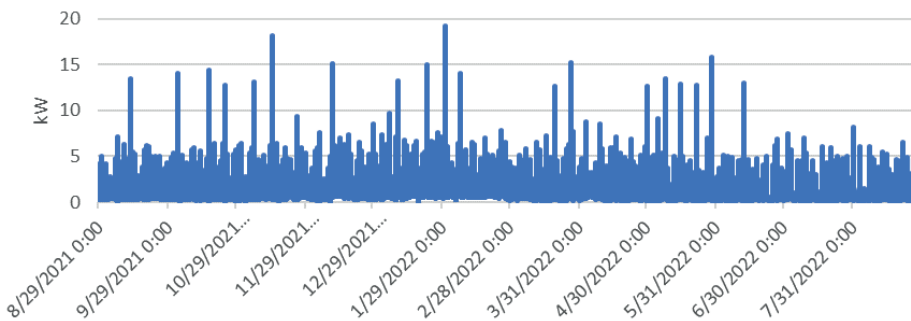


Figure 3. The residential load throughout the year.

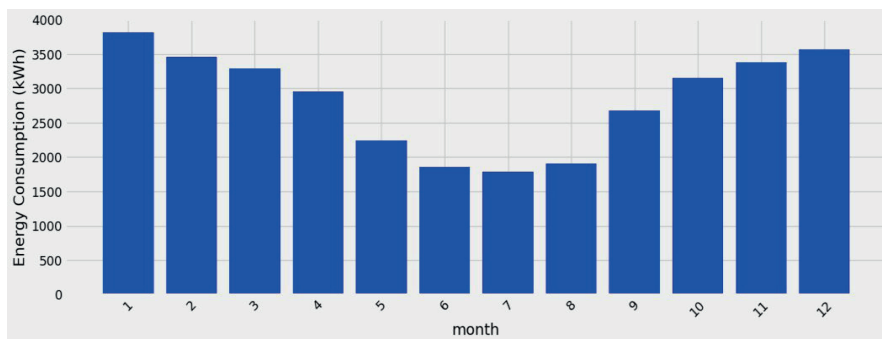


Figure 4. The monthly energy consumption.

The hourly energy consumption data is presented in Figure 5 indicating the highest energy utilization during the evening hours around 7 and 8 p.m., coinciding with most occupants being at home. While there isn't a specific hour of lowest energy consumption evident in the analysis, energy usage tends to be lower in the early morning hours between 2 to 6 a.m. In Figure 6, the individual load patterns of devices like stoves, rainwater drainers, sauna, sockets, water pumps, washing machines, lights and heating are shown during a single day. The sauna, washing machine, electric stove, heater, and water pump are the most energy-consuming loads when they are being utilized.

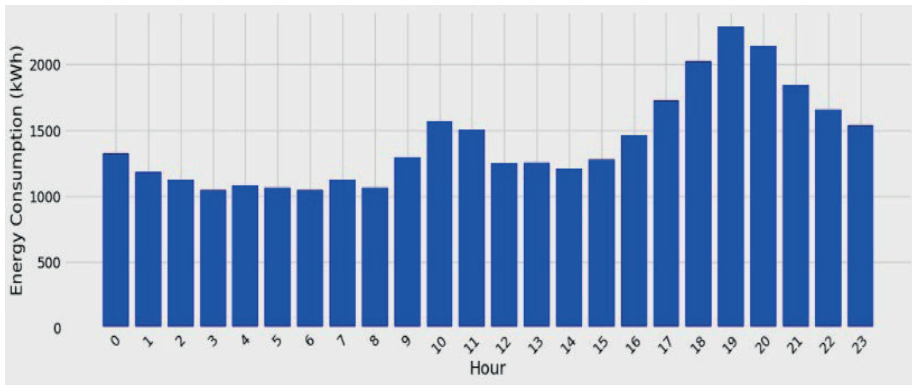


Figure 5. The hourly energy consumption throughout the year.

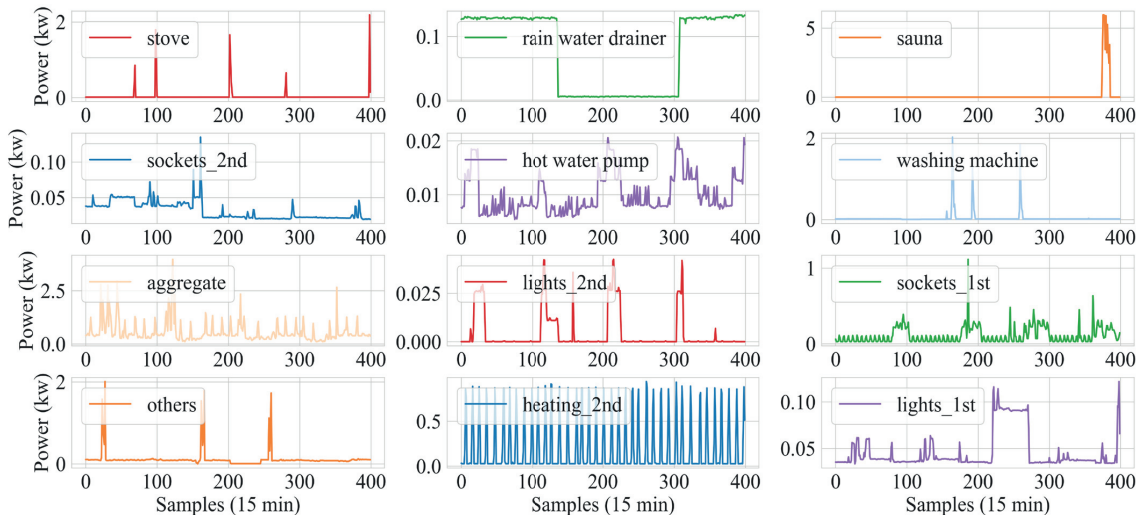


Figure 6. The individual device load pattern in a day.

4.2. Development of NILM based Models

In this research, all compilation has been done utilizing Python 3.10, TensorFlow 2.10.1, and Scikit-Learn 1.4.1, running on a desktop with Intel(R) Core (TM) i7-7700K 4.20 GHz CPU, NVIDIA GeForce GTX 1080 GPU and 32 GB DDR4 RAM. The XgBoost and LSTM models have been trained on a GPU, utilizing the CUDA toolkit version 12.4. However, the Logistic Regression and DTW_KNN models have been trained on a CPU.

Data preprocessing lays the groundwork for effective model training. In the initial data preparation phase, handling missing values is crucial. Mean imputation (replacing missing values with the mean of available data) or forward fill (propagating the last observed value) can be utilized as more general and potentially effective solutions to address missing values. These methods offer more flexibility in handling different types of missing data while preserving the integrity of the dataset. The train-test split, typically at 80/20 ratio ensures unbiased model assessment, additionally shuffling the data during splitting ensures randomness and prevents any inherited order from affecting model performance. It is worth noticing that shuffling the data is not performed when working with LSTM due to the sequential nature of the data. This process is omitted to maintain the integrity of the temporal relationships within the dataset, ensuring optimal performance of the LSTM model. The model specifications for different algorithms are given in Table 2.

Table 2. Model specifications for different algorithms.

Logistic Regression	Solver: "lbfgs", Penalty: "L2", Class Weighted: "balanced", Max_Iteration: 150, Data Shuffle: Yes.
DTW_KNN	Number of Neighbors: 5, Window Size: 150 min (10 samples), Sample Signal Length: 360, Data Shuffle: No.
XgBoost	Booster: "dart", Device: "GPU", eta: 0.5, Max Depth: 4, Min_Child_Depth: 1, Max_Delta_Step: 1, Sub Sample: 0.9, Sampling Method: "Subsample", Objective: "Binary Logistic", Evaluation Metrics: "log loss", Early Stopping: 10, Data Shuffle: Yes.
LSTM	Layer 1: Units: 50, Number of Features: 11 Layer 2: Dense:1, Activation: "Sigmoid" Optimizer: "Adam", Learning Rate: 0.0005, Epochs: 60, Batch size: 32, Early Stopping: 10, metrics: "Accuracy", Loss: "Binary Cross Entropy", Validation Split: 0.1, Class Weight: "Balanced", Data Shuffle = No

Time-related features play a significant role in modeling. Extracting the hour of the day, time interval number (with 15 min resolution), day of the week, and month provides valuable context. Generating cosine and sine values for these above-mentioned features encodes cyclic behavior and enhances models' ability to learn from data. Additionally labeling weekends and holidays provides further insights for predictive modeling. Manual labeling on/off state of appliances based on specific thresholds ensures that the model can learn the underlying patterns as these labels serve as our target variable for supervised learning.

Handling power consumption patterns, particularly for devices with consistent steady consumption, requires a method that effectively identifies meaningful deviations in power usage, filtering out noise and focusing on relevant changes. The approach involves establishing a baseline consumption level for devices such as sockets and lights, representing the minimal power draw when inactive. Significant increases in power consumption beyond this baseline are then interpreted as the device being turned on. Additionally, recognizing that certain devices may exhibit consistent consumption patterns, such as modems, allows for their exclusion to prevent false positives. Overall, this approach balances sensitivity in detecting genuine "on" states with specificity in avoiding false positives, offering a practical means to enhance energy consumption prediction models.

To avoid overfitting in our LSTM and XgBoost networks, we employ early stopping. This method halts the training when the model fails to improve on the validation data after several attempts. It monitors metrics such as loss or accuracy and terminates the training prematurely. This ensures that the model performs well with new data by stopping at the optimal moment. However, we acknowledge that including the training behavior of the models under investigation can offer additional insights into the design process. To this end, we have provided some examples related to the training procedures of LSTM and XgBoost networks below. Note that some of the training sessions ended before reaching the maximum epoch number due to the early stopping callback. Figure 7 shows the logarithmic loss curves related to XgBoost. In Figure 8, the LSTM training and validation losses are depicted.

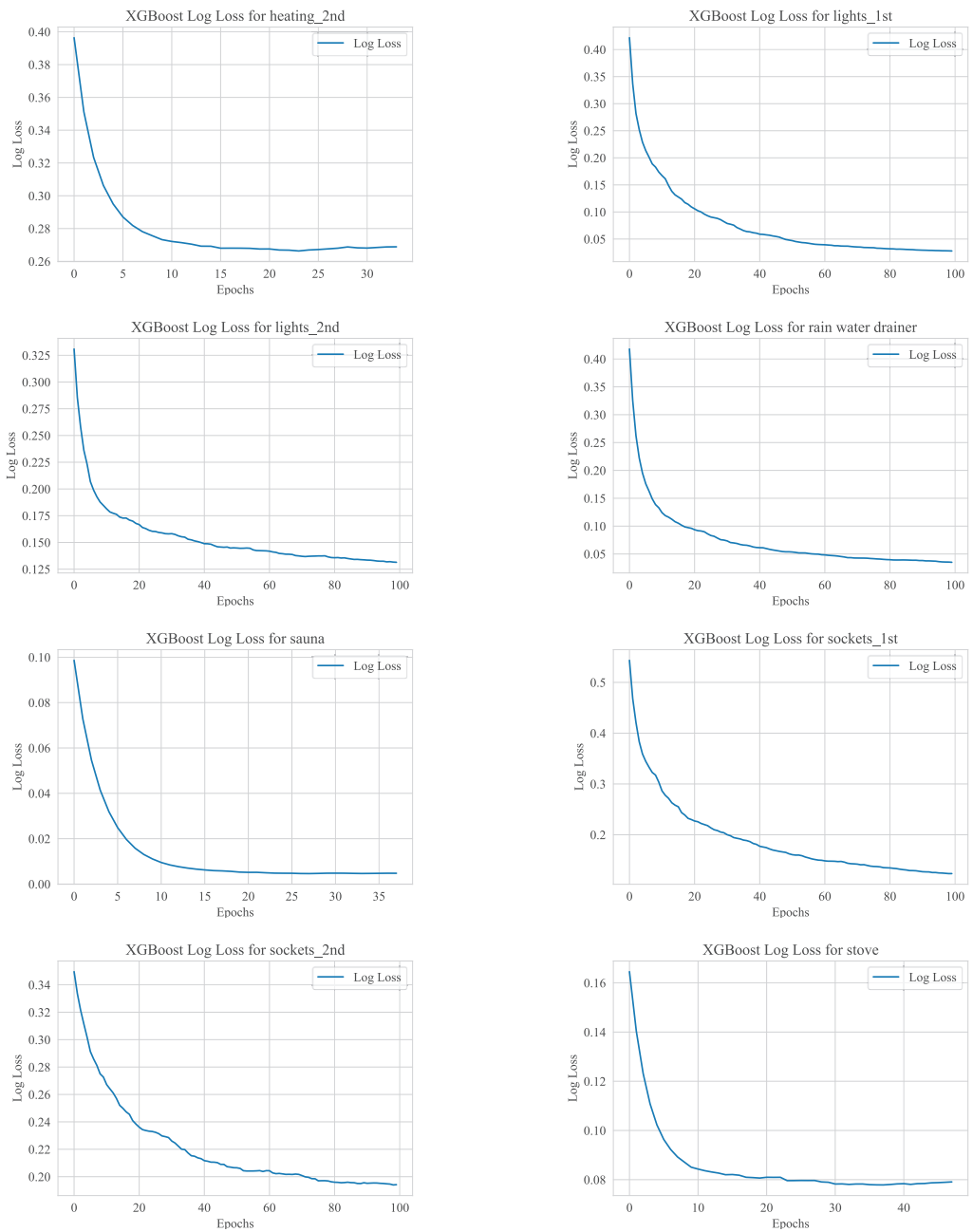


Figure 7. Logarithmic loss curves related to XgBoost training procedure.

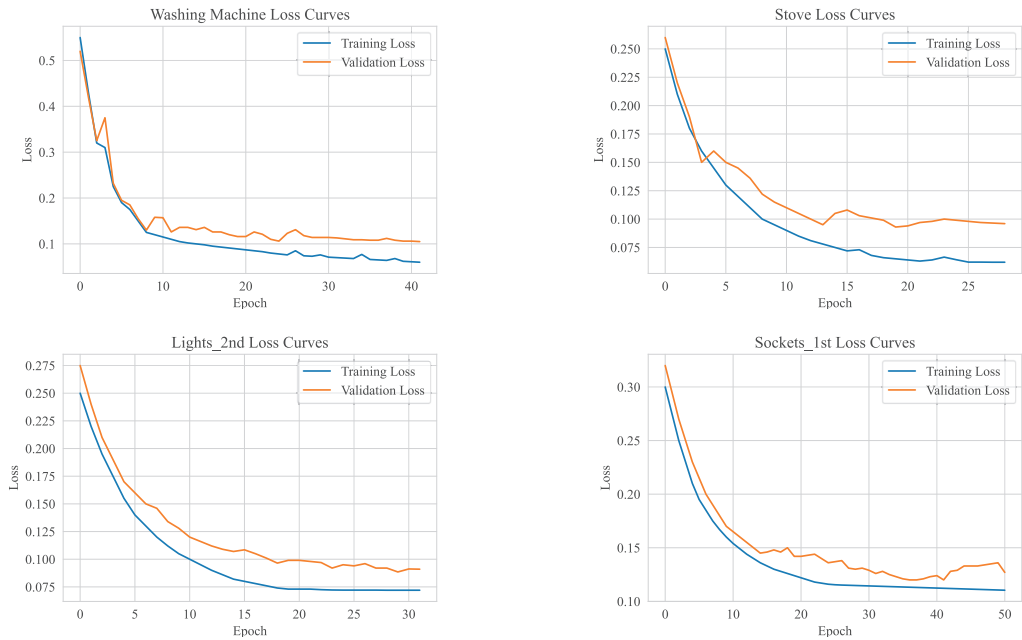


Figure 8. LSTM networks training and validation loss errors for some of the appliances.

5. Results and Discussion

As previously mentioned, our data comprise a 1-year aggregated record of electricity demand, including both the overall demand and the demands and consumption patterns of each appliance, with a resolution of 15 min. Consequently, the dataset encompasses approximately 35,000 data measurements for each sample. We have allocated 80% of the data for training and validation, and 20% for testing purposes. It is critical to highlight that we selected a 1-year period to capture all fluctuations related to seasonality. For example, the sauna is mostly used during the colder seasons, while during summer, the dataset records very few instances of sauna usage. This pattern holds true for heating systems as well.

As you truly mentioned a training time and resource usage investigation is also very important to make a fair comparison among proposed methods. To this end, the training time and RAM resource usage for all the models have been provided in Table 3.

The XgBoost algorithm stands out with exceptional performance across most cases, except for the “other” labeled group, which likely encompasses aggregated power consumption or unknown loads. Given its consistent performance, XgBoost emerges as a robust choice for the given task. The logistic regression demonstrates varying success rates, achieving optimal results in detecting sauna status and rainwater drainers but faltering in other cases.

Dynamic time warping with K-nearest neighbors segments consumption curves into 400-length samples, employing a warping window size of 10 samples to determine appliance classes based on the five nearest neighbors. While generally effective, misidentifications between appliances like washing machines and stoves indicate room for improvement—perhaps through refined feature engineering. Additionally, mislabeling lights_1st and lights_2nd as “others” due to their similar patterns underscores the method’s susceptibility to mixing closely related patterns.

Table 3. Comparison of computation cost training times.

Appliance Name	Method	RAM Usage	Training Time Duration
Stove	LSTM	652 MB	104.62 s
	XgBoost	13.27 MB	2.58 s
	Logistic Regression	17 MB	0.156 s
Sockets_2nd	LSTM	456 MB	228.74 s
	XgBoost	4.6 MB	8.07 s
	Logistic Regression	0.87 MB	0.027 s
Heating_2nd	LSTM	381 MB	141.05 s
	XgBoost	5.65 MB	3.97 s
	Logistic Regression	1.12 MB	0.053 s
Washing Machine	LSTM	506 MB	288.38 s
	XgBoost	43 MB	7.84 s
	Logistic Regression	1.6 MB	0.16 s
Rainwater Drainer	LSTM	602 MB	289.56 s
	XgBoost	13.12 MB	7.31 s
	Logistic Regression	1.15 MB	0.047 s
Sockets_1st	LSTM	925 MB	169.72 s
	XgBoost	60 MB	6.86 s
	Logistic Regression	450 KB	0.05 s
Lights_1st	LSTM	319 MB	160.30 s
	XgBoost	73 MB	7.24 s
	Logistic Regression	626 KB	0.054 s
Sauna	LSTM	113 MB	131.14 s
	XgBoost	27 MB	1.11 s
	Logistic Regression	11 MB	0.027 s
Lights_2nd	LSTM	168 MB	87.52 s
	XgBoost	65 MB	7.37 s
	Logistic Regression	328 KB	0.036 s
Hot Water pump	LSTM	226 MB	165.17 s
	XgBoost	32 MB	2.26 s
	Logistic Regression	10.6 MB	0.16 s

By exploiting the LSTM architecture to analyze 480 samples, representing a window spanning 120 h, the model effectively captures temporal dependencies and inherent patterns within the dataset. The strategic handling of imbalanced labels through the implementation of the class weight method reflects a judicious approach to mitigate bias during model training. The adjustment of class weights ensures equitable consideration of both “on” and “off” states, thereby averting the model’s inclination towards the predominant class and fostering a more balanced learning process.

In the domain of NILM, the classification threshold plays a pivotal role in optimizing the balance between precision and recall. Precision, prioritizing correctness, and risks overlook certain energy consumption patterns, while recall, emphasizing completeness may falsely identify non-existent appliance activations. The choice to adjust the threshold depends on the consequences within the context of energy management. For instance, in residential energy monitoring, minimizing false negatives is crucial to accurately detect

appliance usage, ensuring efficient energy usage and potentially identifying malfunctioning devices. Conversely, in commercial settings like smart buildings, reducing false positives is essential to avoid unnecessary interventions and maintain occupants’ comfort while optimizing energy consumption.

In the provided Figure 9 the confusion matrices indicate that the LSTM method performed less effectively compared to other techniques. One possible reason for this discrepancy could be the suboptimal sampling rate of 15 min. Previous studies have shown that increasing the measurement frequency can significantly enhance prediction accuracy in NILM applications. Additionally, fine-tuning the LSTM method through architectural adjustments or hyperparameter tuning may further improve its performance. Therefore, future research should explore optimizing the sampling frequency alongside other methodological enhancements to maximize the effectiveness of LSTM-based NILM approaches.

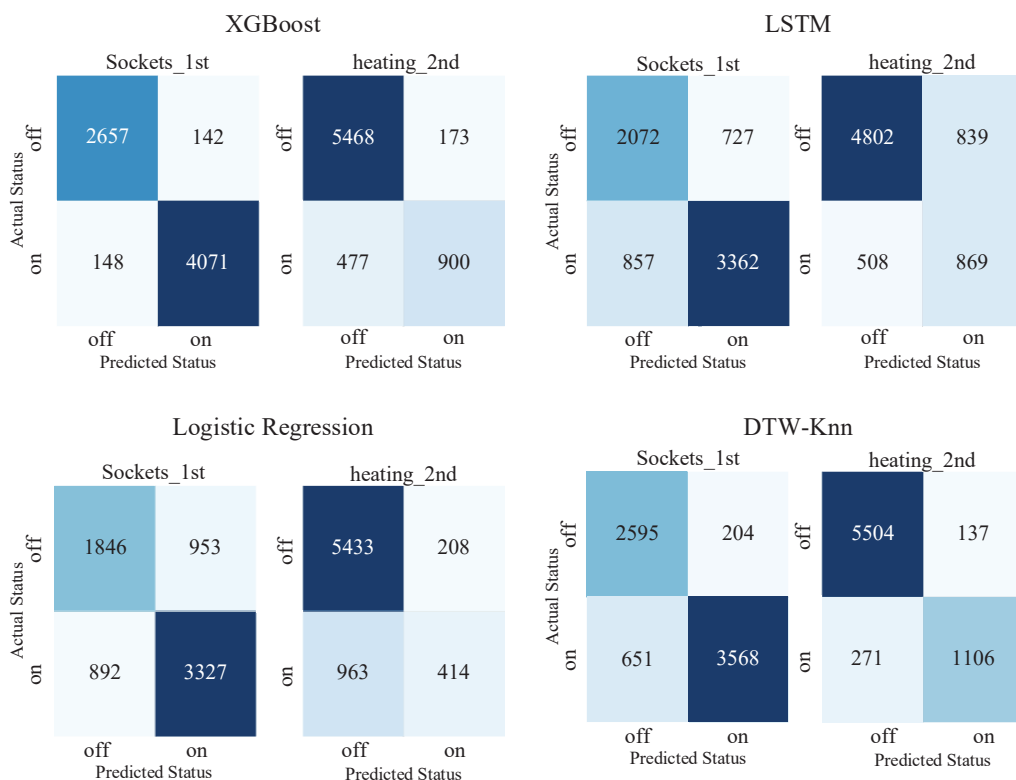


Figure 9. Confusion matrices obtained for specific threshold value.

However, comparative analyses (see Table 4) reveal that alternative models such as XgBoost and DTW with KNN outperform LSTM in the specific scenario under investigation, emphasizing the importance of exploring diverse model architectures and methodologies. Considerations of interpretability, computational efficiency, and ease of implementation will be pivotal in inappropriate model selection for a given task.

Table 4. Performance analysis of different ML algorithms.

Appliance		Stove			Washing Machine			Sauna		
Model Metrics	Precision	Recall	F1-Scroe	Precision	Recall	F1-Scroe	Precision	Recall	F1-Scroe	
DTW_KNN	0.91	0.92	0.95	0.86	0.48	0.62	0.92	0.79	0.85	
XgBoost	0.87	0.73	0.78	0.87	0.74	0.79	0.94	0.93	0.93	
LR	0.71	0.56	0.59	0.72	0.56	0.59	0.95	0.92	0.93	
LSTM	0.63	0.95	0.69	0.58	0.92	0.63	0.76	1.00	0.84	
Appliance		Heating_2nd			Sockets_2nd			Rainwater drainer		
Model Metrics	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	
DTW_KNN	0.91	1.00	0.95	0.94	0.83	0.88	1.00	0.95	0.97	
XgBoost	0.90	0.91	0.90	0.87	0.75	0.79	0.99	0.99	0.99	
LR	0.81	0.83	0.81	0.81	0.88	0.82	0.86	0.86	0.85	
LSTM	0.90	0.89	0.89	0.96	0.92	0.94	0.93	0.96	0.95	
Appliance		Sockets_1st			Lights_1st			Hot water pump		
Model Metrics	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	
DTW_KNN	0.92	0.79	0.85	0.94	0.60	0.73	0.80	0.92	0.86	
XgBoost	0.96	0.96	0.96	0.99	0.98	0.99	0.85	0.88	0.89	
LR	0.74	0.74	0.74	0.77	0.79	0.76	0.70	0.71	0.70	
LSTM	0.83	0.81	0.82	0.82	0.73	0.76	0.60	1.00	0.66	
Appliance		Lights_2nd			Others			Total Performance		
Model Metrics	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	Precision	Recall	F1-scroe	
DTW_KNN	0.90	0.82	0.82	0.52	1.00	0.68	0.93	0.85	0.86	
XgBoost	0.92	0.90	0.91	0.80	0.67	0.71	0.91	0.86	0.88	
LR	0.86	0.88	0.87	0.95	0.97	0.96	0.81	0.79	0.78	
LSTM	0.96	0.86	0.90	0.70	0.81	0.75	0.79	0.90	0.80	

Figure 10 depicts the comparative analysis of all the ML algorithms based on the accuracy of identification of the appliance at the individual level. The accuracy of LSTM and XgBoost is comparable for most of the devices, however, the accuracy of LSTM is extremely low for the lighting loads. On the other side, the LR algorithm has low accurate results for lights, heating, and rainwater drainer. The DTW-KNN algorithm shows comparatively better results than the LR algorithm, but it also has variations in accuracy results. Overall, the most consistent results are from the XgBoost algorithm.

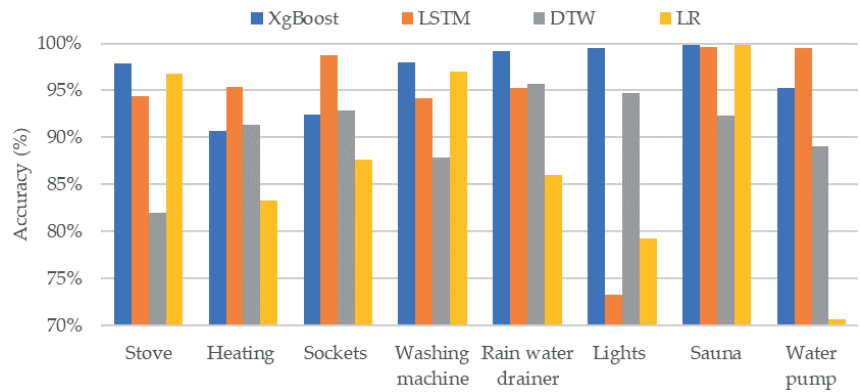


Figure 10. Accuracy analysis of different ML algorithms.

There could be several reasons why XgBoost outperforms other methods. First, XgBoost effectively handles various data types and missing values, which can be beneficial when dealing with different sampling rates. In our case, the sampling rate was quite low.

On the other hand, LSTM can be significantly influenced by the sampling rate. Additionally, the amount of training data has a substantial impact on the performance of neural networks due to their long-term memory capability. Sampling rates matter in Logistic Regression as well. Secondly, XgBoost, as a representative of the family of gradient boosting algorithms, is more robust to overfitting compared to deep learning models like LSTM. This could be particularly beneficial if the data set is not very large, as in our case. XgBoost has many parameters, which gives the designer the opportunity to tune the model and prevent overfitting.

In general, all algorithms have their strengths and weaknesses. For instance, LSTM has the ability to capture long-term dependencies, but it may require a large amount of data and computational resources. Logistic Regression is a simple and fast algorithm, but it may not capture complex patterns in the data. DTW-KNN is good at capturing temporal patterns, but it may be sensitive to noise and outliers. All of the above leads to the conclusion that it is reasonable to focus on developing hybrid models that combine the strengths of different algorithms. Moreover, improving the robustness and efficiency of existing algorithms is valuable as well. This approach not only enhances the performance of the model but also makes it more adaptable to various types of data and tasks.

6. Conclusions

As energy consumption monitoring becomes increasingly vital in the transition towards sustainable practices, this research provides valuable guidance for the selection and deployment of ML techniques in Non-Intrusive Load Monitoring systems. This paper presents a thorough analysis of machine learning techniques employed by NILM through a meticulous examination and comparison, we have elucidated the efficacy and adaptability of various algorithms in disaggregating energy consumption data accurately. Our research underscores the necessity of tailored approaches, emphasizing the significance of selecting suitable models aligned with the specific characteristics and objectives of the data at hand. By providing a nuanced understanding of the strengths and limitations inherent in different methodologies, our study offers valuable insights that can inform the development and implementation of more efficient NILM systems. Furthermore, our findings highlight the multifaceted nature of NILM challenges and the complexity involved in accurately discerning individual appliance signatures from aggregate energy data. The results of this study indicate that the LSTM and XgBoost algorithms give the most accurate identification results, however, XgBoost has the best results on average.

Looking ahead, as the field of NILM continues to evolve, further research and innovation are warranted to address persistent challenges and capitalize on emerging opportunities. By fostering interdisciplinary collaborations and leveraging advances in data science, artificial intelligence, and energy engineering, we can unlock new avenues for improving the accuracy, efficiency, and scalability of NILM solutions. Ultimately, our collective efforts aim to empower consumers with actionable insights, facilitate informed decision-making, and promote sustainable energy consumption practices in support of a more resilient and environmentally conscious future.

Author Contributions: Conceptualization, N.S., K.V. and O.H.; methodology, N.S., K.V. and H.N.H.; software, K.V. and H.N.H.; validation, O.H. and J.B.; formal analysis, N.S., K.V. and O.H.; investigation, N.S. and K.V.; resources, N.S., J.B. and E.P.; data curation, K.V. and H.N.; writing—original draft preparation, N.S., K.V. and H.N.H.; writing—review and editing, O.H., J.B. and E.P.; visualization, K.V. and H.N.H.; supervision, O.H., J.B. and E.P.; project administration, O.H. and E.P.; funding acquisition, O.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Estonian Research Council grant PRG 675 and the European Commission through DUT Horizon Europe Partnership project FLEDGE grant No. MOB3PRT1.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ji, X.; Huang, H.; Chen, D.; Yin, K.; Zuo, Y.; Chen, Z.; Bai, R. A Hybrid Residential Short-Term Load Forecasting Method Using Attention Mechanism and Deep Learning. *Buildings* **2023**, *13*, 72. [\[CrossRef\]](#)
2. Jawad, M.; Asghar, H.; Arshad, J.; Javed, A.; Qureshi, M.B.; Ali, S.M.; Shabbir, N.; Rassolkin, A. A Novel Renewable Powered Stand-Alone Electric Vehicle Parking-Lot Model. *Sustain. Energy Grids Netw.* **2023**, *33*, 100992. [\[CrossRef\]](#)
3. Mirza, Z.T.; Anderson, T.; Seadon, J.; Brent, A. A thematic analysis of the factors that influence the development of a renewable energy policy. *Renew. Energy Focus* **2024**, *49*, 100562. [\[CrossRef\]](#)
4. Abumohsen, M.; Owda, A.Y.; Owda, M. Electrical Load Forecasting Using LSTM, GRU, and RNN Algorithms. *Energies* **2023**, *16*, 2283. [\[CrossRef\]](#)
5. Azizi, E.; Ahmadihangar, R.; Rosin, A.; Bolouki, S. Characterizing Energy Flexibility of Buildings with Electric Vehicles and Shiftable Appliances on Single Building Level and Aggregated Level. *Sustain. Cities Soc.* **2022**, *84*, 103999. [\[CrossRef\]](#)
6. Azizi, E.; Shotorbani, A.M.; Hamidi-Beheshti, M.T.; Mohammadi-Ivatloo, B.; Bolouki, S. Residential Household Non-Intrusive Load Monitoring via Smart Event-Based Optimization. *IEEE Trans. Consum. Electron.* **2020**, *66*, 233–241. [\[CrossRef\]](#)
7. Opoku, R.; Obeng, G.Y.; Adjei, E.A.; Davis, F.; Akuffo, F.O. Integrated System Efficiency in Reducing Redundancy and Promoting Residential Renewable Energy in Countries without Net-Metering: A Case Study of a SHS in Ghana. *Renew. Energy* **2020**, *155*, 65–78. [\[CrossRef\]](#)
8. Azizi, E.; Beheshti, M.T.H.; Bolouki, S. Event Matching Classification Method for Non-Intrusive Load Monitoring. *Sustainability* **2021**, *13*, 693. [\[CrossRef\]](#)
9. Lizana, J.; Friedrich, D.; Renaldi, R.; Chacartegui, R. Energy Flexible Building through Smart Demand-Side Management and Latent Heat Storage. *Appl. Energy* **2018**, *230*, 471–485. [\[CrossRef\]](#)
10. Kaselemi, M.; Protopapadakis, E.; Voulodimos, A.; Doulami, N.; Doulami, A. Multi-Channel Recurrent Convolutional Neural Networks for Energy Disaggregation. *IEEE Access* **2019**, *7*, 81047–81056. [\[CrossRef\]](#)
11. Schirmer, P.A.; Mporas, I. Non-Intrusive Load Monitoring: A Review. *IEEE Trans. Smart Grid* **2023**, *14*, 769–784. [\[CrossRef\]](#)
12. Angelis, G.F.; Timplalexis, C.; Salamanis, A.I.; Krinidis, S.; Ioannidis, D.; Kehagias, D.; Tzovaras, D. Energformer: A New Transformer Model for Energy Disaggregation. *IEEE Trans. Consum. Electron.* **2023**, *69*, 308–320. [\[CrossRef\]](#)
13. Azizi, E.; Beheshti, M.T.H.; Bolouki, S. Quantification of Disaggregation Difficulty with Respect to the Number of Smart Meters. *IEEE Trans. Smart Grid* **2022**, *13*, 516–525. [\[CrossRef\]](#)
14. Shabbir, N.; Kutt, L.; Jawad, M.; Iqbal, M.N.; Ghahfarokhi, P.S. Forecasting of Energy Consumption and Production Using Recurrent Neural Networks. *Adv. Electr. Electron. Eng.* **2020**, *18*, 190–197. [\[CrossRef\]](#)
15. Ding, G.; Wu, C.; Wang, Y.; Liang, Y.; Jiang, X.; Li, X. A Novel Non-Intrusive Load Monitoring Method Based on Quantum Particle Swarm Optimization Algorithm. In Proceedings of the 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), Qiqihar, China, 28–29 April 2019; pp. 230–234. [\[CrossRef\]](#)
16. Nuran, A.S.; Murti, M.A.; Suratman, F.Y. Non-Intrusive Load Monitoring Method for Appliance Identification Using Random Forest Algorithm. In Proceedings of the 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–11 March 2023; pp. 754–758. [\[CrossRef\]](#)
17. Raiker, G.A.; Reddy, B.S.; Umanand, L.; Agrawal, S.; Thakur, A.S.; Ashwin, K.; Barton, J.P.; Thomson, M. Internet of Things Based Demand Side Energy Management System Using Non-Intrusive Load Monitoring. In Proceedings of the 2020 IEEE International Conference on Power Electronics, Smart Grid and Renewable Energy (PESGRE2020), Cochin, India, 2–4 January 2020; pp. 5–9. [\[CrossRef\]](#)
18. Al-Khadher, O.; Mukhtaruddin, A.; Hashim, F.R.; Azizan, M.M.; Mamat, H.; Mani, M. Comparison of Non-Intrusive Load Monitoring Supervised Methods Using Harmonics as Feature. In Proceedings of the 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET), Prague, Czech Republic, 20–22 July 2022; pp. 1–6. [\[CrossRef\]](#)
19. Moradzadeh, A.; Zeinal-Kheiri, S.; Mohammadi-Ivatloo, B.; Abapour, M.; Anvari-Moghaddam, A. Support Vector Machine-Assisted Improvement Residential Load Disaggregation. In Proceedings of the 2020 28th Iranian conference on electrical engineering (ICEE), Tabriz, Iran, 4–6 August 2020; pp. 1–6. [\[CrossRef\]](#)
20. Shiddieqy, H.A.; Hariadi, F.I.; Adijarto, W. Plug-Load Classification Based on CNN from V-I Trajectory Image Using STM32. In Proceedings of the 2021 International Symposium on Electronics and Smart Devices (ISESD), Bandung, Indonesia, 29–30 June 2021. [\[CrossRef\]](#)
21. Giannuzzo, L.; Minuto, F.D.; Schiera, D.S.; Lanzini, A. Reconstructing Hourly Residential Electrical Load Profiles for Renewable Energy Communities Using Non-Intrusive Machine Learning Techniques. *Energy AI* **2024**, *15*, 100329. [\[CrossRef\]](#)
22. Hernández, Á.; Nieto, R.; de Diego-Otón, L.; Pérez-Rubio, M.C.; Villadangos-Carrizo, J.M.; Pizarro, D.; Ureña, J. Detection of Anomalies in Daily Activities Using Data from Smart Meters. *Sensors* **2024**, *24*, 515. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Koasidis, K.; Marinakis, V.; Doukas, H.; Karamaneas, A.; Nikas, A.; Doumouras, N. Equipment- and Time-Constrained Data Acquisition Protocol for Non-Intrusive Appliance Load Monitoring. *Energies* **2023**, *16*, 7315. [\[CrossRef\]](#)
24. Hosseini, S.S.; Delcroix, B.; Henao, N.; Agbossou, K.; Kelouwani, S. Towards Feasible Solutions for Load Monitoring in Quebec Residences. *Sensors* **2023**, *23*, 7288. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Shareef, H.; Asna, M.; Errouissi, R.; Prasanthi, A. Rule-Based Non-Intrusive Load Monitoring Using Steady-State Current Waveform Features. *Sensors* **2023**, *23*, 6926. [\[CrossRef\]](#)

26. Azizi, E.; Beheshti, M.T.H.; Bolouki, S. Appliance-Level Anomaly Detection in Nonintrusive Load Monitoring via Power Consumption-Based Feature Analysis. *IEEE Trans. Consum. Electron.* **2021**, *67*, 363–371. [[CrossRef](#)]
27. Welikala, S.; Dinesh, C.; Ekanayake, M.P.B.; Godaliyadda, R.I.; Ekanayake, J. Incorporating Appliance Usage Patterns for Non-Intrusive Load Monitoring and Load Forecasting. *IEEE Trans. Smart Grid* **2019**, *10*, 448–461. [[CrossRef](#)]
28. Tao, Y.; Qiu, J.; Lai, S.; Wang, Y.; Sun, X. Reserve Evaluation and Energy Management of Micro-Grids in Joint Electricity Markets Based on Non-Intrusive Load Monitoring. *IEEE Trans. Ind. Appl.* **2023**, *59*, 207–219. [[CrossRef](#)]
29. Ma, C.; Yin, L. Deep Flexible Transmitter Networks for Non-Intrusive Load Monitoring of Power Distribution Networks. *IEEE Access* **2021**, *9*, 107424–107436. [[CrossRef](#)]
30. Garcia, F.D.; Souza, W.A.; Diniz, I.S.; Marafão, F.P. NILM-Based Approach for Energy Efficiency Assessment of Household Appliances. *Energy Inform.* **2020**, *3*, 10. [[CrossRef](#)]
31. Liu, H.; Wu, H.; Yu, C. A Hybrid Model for Appliance Classification Based on Time Series Features. *Energy Build.* **2019**, *196*, 112–123. [[CrossRef](#)]
32. Gopinath, R.; Kumar, M.; Prakash Chandra Joshua, C.; Srinivas, K. Energy Management Using Non-Intrusive Load Monitoring Techniques—State-of-the-Art and Future Research Directions. *Sustain. Cities Soc.* **2020**, *62*, 102411. [[CrossRef](#)]
33. Du, Z.; Yin, B.; Zhu, Y.; Huang, X.; Xu, J. A NILM Load Identification Method Based on Structured V-I Mapping. *Sci. Rep.* **2023**, *13*, 21276. [[CrossRef](#)]
34. Reddy, R.; Garg, V.; Pudi, V. A Feature Fusion Technique for Improved Non-Intrusive Load Monitoring. *Energy Inform.* **2020**, *3*, 9. [[CrossRef](#)]
35. Ruano, A.; Hernandez, A.; Ureña, J.; Ruano, M.; Garcia, J. NILM Techniques for Intelligent Home Energy Management and Ambient Assisted Living: A Review. *Energies* **2019**, *12*, 2203. [[CrossRef](#)]
36. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [[CrossRef](#)]
37. He, K.; Stankovic, V.; Stankovic, L. Building a Graph Signal Processing Model Using Dynamic Time Warping for Load Disaggregation. *Sensors* **2020**, *20*, 6628. [[CrossRef](#)]
38. Shabbir, N.; Ahmadihangar, R.; Rosin, A.; Astapov, V.; Kilter, J. Short-Term PV Energy Generation Forecasting Using Deep Learning. *Telematique* **2022**, *21*, 5844–5850.
39. Wang, R.; Tao, D. DTW-KNN Implementation for Touch-Based Authentication System. In Proceedings of the 2019 5th International Conference on Big Data Computing and Communications (BIGCOM), QingDao, China, 9–11 August 2019; pp. 318–322. [[CrossRef](#)]
40. Feng, Z.; Wang, M.; He, J.; Xiao, W. Real-Time Equipment State Monitoring and Cycle Time Calculation Method Based on DTW-KNN. In Proceedings of the 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 21–23 April 2023; pp. 1350–1353. [[CrossRef](#)]
41. Angelis, G.F.; Timplalexis, C.; Krinidis, S.; Ioannidis, D.; Tzovaras, D. NILM Applications: Literature Review of Learning Approaches, Recent Developments and Challenges. *Energy Build.* **2022**, *261*, 111951. [[CrossRef](#)]
42. Imtiaz, M.A.; Raja, G. Isolated Word Automatic Speech Recognition (ASR) System Using MFCC, DTW & KNN. In Proceedings of the 2016 Asia Pacific Conference on Multimedia and Broadcasting (APMediaCast), Bali, Indonesia, 17–19 November 2017; pp. 106–110. [[CrossRef](#)]
43. Yang, N.C.; Sung, K.L. Non-Intrusive Load Classification and Recognition Using Soft-Voting Ensemble Learning Algorithm with Decision Tree, K-Nearest Neighbor Algorithm and Multilayer Perceptron. *IEEE Access* **2023**, *11*, 94506–94520. [[CrossRef](#)]
44. Shabbir, N.; Ahmadihangar, R.; Rosin, A.; Jawad, M.; Kilter, J.; Martins, J. XgBoost Based Short-Term Electrical Load Forecasting Considering Trends & Periodicity in Historical Data. In Proceedings of the 2023 IEEE International Conference on Energy Technologies for Future Grids (ETFG), Wollongong, Australia, 3–6 December 2023; pp. 1–6. [[CrossRef](#)]
45. Chen, Z.; Chen, J.; Xu, X.; Peng, S.; Xiao, J.; Qiao, H. Non-Intrusive Load Monitoring Based on Feature Extraction of Change-Point and Xgboost Classifier. In Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 30 October–November 2020; pp. 2652–2656. [[CrossRef](#)]
46. Semmelmann, L.; Henni, S.; Weinhardt, C. Load Forecasting for Energy Communities: A Novel LSTM-XGBoost Hybrid Model Based on Smart Meter Data. *Energy Inform.* **2022**, *5*, 24. [[CrossRef](#)]
47. Mo, H.; Sun, H.; Liu, J.; Wei, S. Developing Window Behavior Models for Residential Buildings Using XGBoost Algorithm. *Energy Build.* **2019**, *205*, 109564. [[CrossRef](#)]
48. Bipin Nair, B.J.; Yadhukrishnan, S.; Manish, A. A Comparative Study on Document Images Classification Using Logistic Regression and Multiple Linear Regressions. In Proceedings of the 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 23–25 August 2023; pp. 1096–1104. [[CrossRef](#)]
49. Shabbir, N.; Kütt, L.; Jawad, M.; Husev, O.; Rehman, A.U.; Gardezi, A.A.; Shafiq, M.; Choi, J.G. Short-Term Wind Energy Forecasting Using Deep Learning-Based Predictive Analytics. *Comput. Mater. Contin.* **2022**, *72*, 1017–1033. [[CrossRef](#)]
50. Shabbir, N.; Kütt, L.; Raja, H.A.; Ahmadihangar, R.; Rosin, A.; Husev, O. Machine Learning and Deep Learning Techniques for Residential Load Forecasting: A Comparative Analysis. In Proceedings of the IEEE 62nd International Scientific Conference on Power and Electrical Engineering of Riga Technical University, Riga, Latvia, 15–17 November 2021; pp. 1–5.
51. Mellit, A.; Pavan, A.M.; Lughii, V. Deep Learning Neural Networks for Short-Term Photovoltaic Power Forecasting. *Renew. Energy* **2021**, *172*, 276–288. [[CrossRef](#)]

52. Hossen, T.; Nair, A.S.; Chinnathambi, R.A.; Ranganathan, P. Residential Load Forecasting Using Deep Neural Networks (DNN). In Proceedings of the 2018 North American Power Symposium (NAPS), Fargo, ND, USA, 9–11 September 2019. [[CrossRef](#)]
53. Shabbir, N.; Ahmadiyahangar, R.; Rosin, A.; Husev, O.; Jalakas, T.; Martins, J. Residential DC Load Forecasting Using Long Short-Term Memory Network (LSTM). In Proceedings of the 2023 IEEE 11th International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 13–15 August 2023; pp. 131–136. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Publication IV

H. N. Hokmabad, T. H. Shahsavari, O. Matiushkin, T. Jalakas, O. Husev, and J. Belikov, 'Single Cell Energy Router Justification for Three Phase Near Zero Energy Buildings', in *2025 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2025, pp. 1622–1628.

Single Cell Energy Router Justification for Three Phase Near Zero Energy Buildings

Hossein Nouroollahi Hokmabad
Department of Electrical Power
Engineering and Mechatronics
Tallinn University of Technology
Tallinn, Estonia
hossein.nouroollahi@taltech.ee

Tanel Jalakas
Department of Electrical Power
Engineering and Mechatronics
Tallinn University of Technology
Tallinn, Estonia
tanel.jalakas@taltech.ee

Tala Hemmati Shahsavari
Department of Electrical Power
Engineering and Mechatronics
Tallinn University of Technology
Tallinn, Estonia
talahemmati@gmail.com

Oleksandr Husev
Faculty of Electrical and Automation
Engineering
Gdansk University of Technology
Gdansk, Poland
oleksandr.husev@gmail.com

Oleksandr Matiushkin
Department of Electrical Power
Engineering and Mechatronics
Tallinn University of Technology
Tallinn, Estonia
oleksandr.matiushkin@taltech.ee

Juri Belikov
Department of Software Science
Tallinn University of Technology
Tallinn, Estonia
juri.belikov@taltech.ee

Abstract—The rise of building-integrated photovoltaics and distributed electric vehicle charging has led to significant phase imbalances in utility grids, challenging service providers due to limited behind-the-meter visibility. This paper introduces a novel Single-Cell Three-Phase (SC-TP) Energy Router (ER) that accesses all phases and balances them without the complexities and costs associated with conventional three-phase systems. Our comparative analysis shows that the SC-TP ER reduces phase unbalancing by 16% and achieves cost savings, offering a viable alternative to three-phase solutions with 25% reduction in cost.

Keywords—near zero energy building, renewable energies, grid congestion, phase imbalance, phase balancing, smart grids

I. INTRODUCTION

In recent years, the rapid integration of renewable energy sources, especially Photovoltaic (PV) systems, has significantly reshaped the topology of modern power distribution networks. This shift requires power systems to accommodate an increasing share of intermittent renewable generation while addressing new complexities in managing localized and distributed energy sources. Grid congestion and phase imbalances are intensifying due to the rise of small-scale renewable setups and Electric Vehicle (EV) charging stations, particularly in residential sectors. Additionally, phenomena like the duck curve are challenging the reliability of electricity grids [1], [2].

Phase imbalances reduce grid reliability and resiliency by decreasing power capacity and increasing losses in transformer secondaries [3], [4]. Expanding feeder capacity to address these issues would impose significant costs for Distribution System Operators (DSOs) [5], [6]. Consequently, alternative solutions which do not require large investments are preferred. One proposed solution involves clustering residential prosumers equipped with PV-Energy Storage (PV-ES) systems and high-demand buildings into evenly distributed three-phase microgrids. However, this approach faces challenges due to limited visibility of behind-the-meter installations [7], [8], [9].

Enhancing the intelligence of low-voltage equipment and integrating advanced power electronic technologies provide

innovative and cost-effective solutions for phase balancing, potentially reducing or delaying the need for costly grid reinforcement [10], [11]. To this end, various solutions have been proposed. For example, some researchers have developed data-driven approaches for identifying and mitigating phase imbalances [12], [13], [14]. Others have introduced a two-stage network balancing strategy that employs phase-switch devices at terminal nodes and phase-switching soft open points across grid nodes [15]. Additionally, phase balancing challenges may also arise in highly Inverter-Based Resource (IBR) penetrated grids, where IBRs must form and stabilize the grid in the absence of a central generator. To address these scenarios, [16] proposes a solution in which inverters use droop control mechanisms to balance phases within the distribution network.

Demand-side phase balancing, proposed as a complementary solution to dispatch-side balancing in [17], involves regulating asymmetric loads. This approach integrates an optimization model that manages load imbalances from the demand side while coordinating with step voltage regulators and dispatching distributed generators to enhance overall phase stability and balance. However, with effective management and coordination, these high-demand loads could be leveraged as grid-balancing resources, offering support to DSOs [18]. Yet, there are currently very few market-ready options that autonomously address these issues without requiring end-user involvement. A promising alternative is presented in [19], which proposes an incentive-based scheme encouraging flexible consumers to assist in phase balancing. This method employs a centralized control algorithm that utilizes customers' installed converters for balancing phases at the substation. However, the paper assumes infrastructure readiness without addressing its practical implementation or scalability challenges.

While extensive research exists, practical solutions for grid phase balancing with PV, ES, and EVs often overlook the economic feasibility of the developed systems, making them less attractive to the market. This paper bridges the gap between practicality and economic efficiency by proposing a novel topology that detects unbalanced phases at terminal points and selectively directs PV or ES power to equalize

interactions across each phase with the three-phase electricity grid. The proposed single-cell topology connects to all phases and includes an integrated non-isolated 350 V and isolated 48 - 350 V direct current (dc) bus alongside alternating current (ac) lines, enabling direct power supply to dc loads. This design reduces unnecessary energy conversion losses and minimizes harmonic injection into the ac grid.

II. PROPOSED SINGLE CELL ENERGY ROUTER CONCEPT

A. Structure of Single Cell ER:

Fig. 1 shows the proposed topology for a Single-Cell Three-Phase (SC-TP) Energy Router (ER) and its experimental realization. In this configuration, a dc bus can interact with all ac phases through an ER, but not simultaneously. Phase balance can be enhanced by detecting and reducing the power consumption of the phase with the highest demand. Since the power drawn from the three phases in a three-phase connected buildings often varies significantly, reducing the phase imbalance ratio with the proposed solution could lead to substantial economic benefits by eliminating the need for two additional converting cells. This advantage supports the rationale for adopting a single-cell approach rather than a conventional three-phase system. To achieve this goal, a Smart Energy Management Algorithm (SEMA) running beside essential low level controlling algorithm is needed to enable the ER to detect, mitigate and smooth the phase differences. Accordingly, the method proposed in [20] has been selected as the SEMA, and details regarding the energy flow optimization can be reviewed there. The technical characteristic of ER is collected in Table I. Further details related to the SC-SP ER topology is introduced in [21]. An alternative approach, which also offers dual purpose applications for both ac and dc grid is proposed in [22].

B. Test Study and Analyzed Scenarios:

To validate our hypothesis and assess the effectiveness of the proposed topology, we collected and analyzed annual load consumption data from a residential house located in Tallinn, Estonia, where the owner has an EV, and the dwelling is

connected to the electricity grid via a three-phase terminal. The data were collected at a 3-second resolution; however, for simplicity and improved visualization, it was averaged and down sampled to a 1-hour resolution. Fig. 2, illustrates and compares a snapshot of initial phase imbalances and phase statuses after ideal phase balancing actions. In initial mode, it is evident that phase 3, labeled "L3", delivers less power compared to the other two phases, indicating that the EV charger should be connected to this phase. However, as observed, severe phase imbalance occurs during EV charging, regardless of which phase the EV is linked.

In this study, three ER topologies include: Single-Cell Single-Phase (SC-SP), Three-Cell Three-Phase (TC-TP), and SC-TP, are considered and their capability for phase balancing have been compared with each other. The topologies performances are first simulated and then validated with an experimental setup. The SC-TP ER is the upgraded version of the SC-SP ER. The 8 kWh Li-Ion battery pack is used as an ES. The methodology for selecting the optimal ES capacity is demonstrated in [23].

TABLE I. ER TECHNICAL PARAMETERS

Parameters	Value
Rated power	15 kW
Grid and load side ac voltage (RMS)	230 V – 50 HZ
dc-link voltage	350 V
Nominal current of each phase	25 A
Switching frequency	65 kHz
Solar voltage input range	150 - 600 V
ES voltage input range	150 – 330 V
dc-link capacitor	3 mF

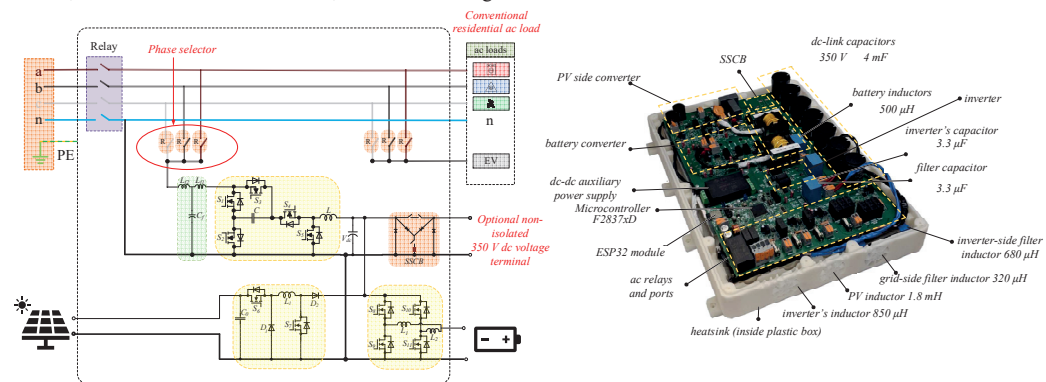


Fig. 1. Abstract view of the proposed single cell topology connected to three-phase terminal with EV charger, PV, and ES integration and its experimental realization.

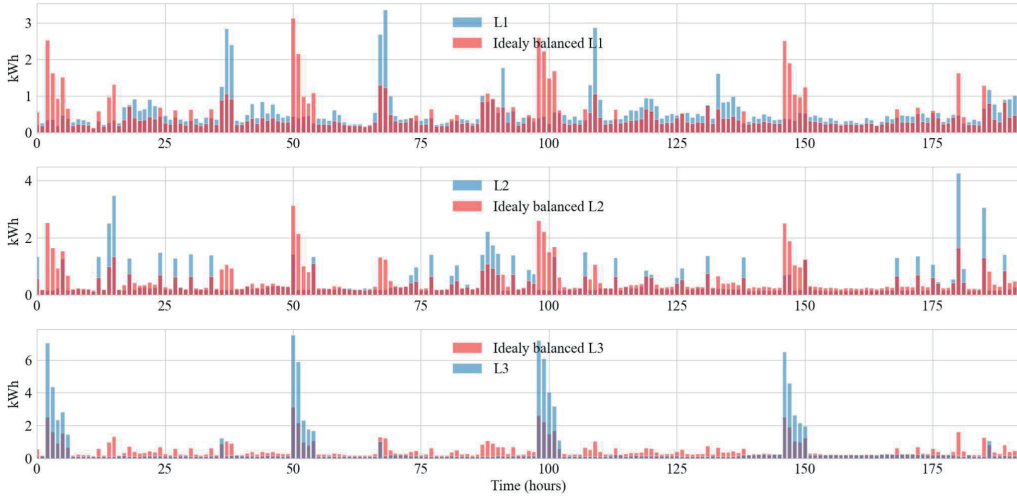


Fig. 2. Imbalanced and balanced phase power range comparison assuming ideal phase balancing possibilities.

C. Phase imbalance:

In this study, the phase imbalance ratio is defined as:

$$UB \% = \sum_{t=0}^T \frac{L_{unbl,t}}{L_{avg,t}} \times \frac{100}{N} \quad (1)$$

$$L_{unbl,t} =$$

$$\text{Max} (|L_{avg,t} - L_{1,t}| + |L_{avg,t} - L_{2,t}| + |L_{avg,t} - L_{3,t}|) \quad (2)$$

$$L_{avg,t} = (L_{1,t} + L_{2,t} + L_{3,t})/3 \quad (3)$$

where $L_{1,t}, L_{2,t}, L_{3,t}$ is a power consumption for phase 1, 2, 3 at time t , respectively. $L_{avg,t}$ is the average power consumption from all phases in time t and $L_{unbl,t}$ is a maximum power deviation from $L_{avg,t}$ at time t . N is the total number of time steps in which the phase unbalance is calculated. Finally, $UB \%$ is the average phase imbalance. In order to minimize the $UB \%$ ratio an optimization problem is formulated for the TC-TP mode as:

$$\text{Min } f(x) = \text{Min} (|L_{avg}'' - L_1''| + |L_{avg}'' - L_2''| + |L_{avg}'' - L_3''|) \quad (4)$$

subject to:

$$L_1'', L_2'', L_3'' \geq 0 \quad (5)$$

$$PV_{l_1} + PV_{l_2} + PV_{l_3} \leq PV \quad (6)$$

$$ES_{l_1} + ES_{l_2} + ES_{l_3} \leq ES \quad (7)$$

$$L_i - PV_{l_i} - ES_{l_i} = L_i'', \quad i = 1, 2, 3 \quad (8)$$

where $L_{avg}'' = (L_1'' + L_2'' + L_3'')/3$ represents the average demand and L_y'' is each phase's demand from the electricity grid after allocating available renewable energy resources. Here, y denotes the phase number. PV_{l_y} and ES_{l_y} represents the allocated energy from solar energy production and energy storage, respectively. PV and ES denote the total accessible energy from renewable setups and batteries, respectively. In Eq. (4), x represents optimization factors, $x = [PV_{l_1}, PV_{l_2}, PV_{l_3}, ES_{l_1}, ES_{l_2}, ES_{l_3}]$. Finally, the ES SoC level should be updated as:

$$ES_{t+1} = \min(ES_{max}, ES - ES_{l_1} - ES_{l_2} - ES_{l_3} + PV') \quad (9)$$

where, ES_{t+1} is the ES, State of Charge (SoC) (%) for the next time step, and ES_{max} denotes the maximum energy capacity of ES, and PV'_t denotes the remaining generated solar energy after demand responding. It should be mentioned that, in all equations, we are considering energy instead of power, since this assumption simplifies the equations.

In SC-TP mode, the optimization problem must be reformulated. Algorithm 1 outlines the optimization process for SC-TP mode, where the phase with the maximum load demand is identified at each time step. The ER then links this phase to the DC link and injects renewable or stored energy to meet the demand, in the selected phase. In SC-SP mode, since phase exchange is not possible and the ER remains permanently connected to phase "L1," the optimization problem simplifies into a reduced version of Algorithm 1, with the index consistently set to 1. The optimization problems are solved using the Pyomo interface with the IPOPT solver, due to the nonlinear nature of the formulation. All code related to optimization process is implemented in Python 3.11.

Algorithm 1 SC-TP Energy Optimization Algorithm

Input:

- 1: Annual load profiles ($\vec{l}_1, \vec{l}_2, \vec{l}_3$), Annual PV profile (PV).
- 2: **for** $i = 1$ to n **do**:
- 3: $l_{index}, index = \max(l_{1,i}, l_{2,i}, l_{3,i})$
- 4: $Index = [1, 2, 3]$
- 5: $x, y = Index.drop(index)$
- 6: $Min f(x) = \min(|l_{avg,i}^n - l_{index,i}^n| + |l_{avg,i}^n - l_{x,i}^n| + |l_{avg,i}^n - l_{y,i}^n|)$
- 7: Subject to:
- 8: $l_{avg,i}^n = avg(l_{index,i}^n, l_{x,i}^n, l_{y,i}^n)$
- 10: $l_{index,i}^n \geq 0$
- 11: $l_{index,i}^n = l_{index,i} - PV_i - ES_i^n$
- 12: $PV_i^n \leq PV_i$
- 13: $ES_i^n \leq ES_i$
- 14: $ES_{i+1} = \min(ES_{max}, ES_i - ES_i^n + PV_i - PV_i^n)$
- 22: **end for**
- 23: **Return:** optimal values PV_i^n, ES_i^n

III. CASE STUDIES EVALUATION AND DISCUSSION

In this section, the outcomes of the proposed methodology are discussed from multiple perspectives, including technical performance, operational feasibility, and economic impact. The phase balancing potential of all possible topologies is compared, followed by a discussion of the experimental realization of the SC-TP ER. Finally, a cost comparison is provided to highlight the impact of the proposed topology on cost reductions and its potential for practical adoption in building with three-phase grid-connected systems.

A. Phase Balancing Outcomes:

Fig. 3 presents a normalized radar chart comparison across all topologies, measuring variables such as UB%, PV self-consumption ratio, average SoC (%) of ES, capital costs and average load per phase during an experimental test period. Each category is normalized to its maximum observed value. For example, the average load per phase is highest when the house does not utilize any local renewable energy sources.

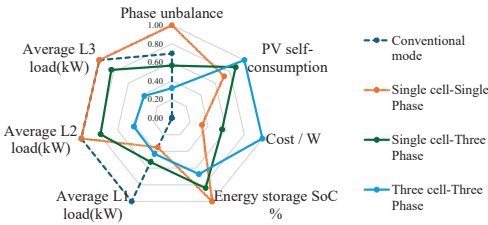


Fig. 3. Normalized radar chart comparison for all topologies.

In conventional houses, the phase balance ratio is better than in scenarios where the PV setup is connected to only one phase. The greatest phase imbalance is observed in the SC-SP topology, indicating that the presence of PV in this configuration reduces the phase balance ratio. This is because, in such operational modes, the generated renewable energy is only injected into one phase, significantly lowering the demand on that phase alone. Meanwhile, the other phases continue to draw the same demand from the grid, unable to benefit from PV production or stored energy. Considering that the SC-SP topology is a commonly used option in residential

PV setups due to its affordability [24], this underscores the necessity of proposing novel solutions for improving phase balance ratio.

The TC-TP topology outperforms other configurations in all categories, including PV self-consumption ratio, phase balance ratios, and average ES's SoC (%) levels. However, its capital cost ratio is 1.35 times higher than that of the SC-TP topology, posing a further barrier to adoption in residential buildings, where economic feasibility is a critical factor for end-users. Furthermore, the TC-TP topology can reduce grid interaction across all phases at a similar rate, whereas the SC-TP topology does not achieve the same performance, with the management algorithm primarily focused on reducing interactions on phase L1. This is because, for most of the time, the demand on L1 is higher than on the other two phases, prompting the ER to link PV and ES resources to this phase. It is important to note that phase switching frequency is constrained by numerous factors; in this study, it is set to 15-minute intervals. Improving phase-switching algorithms and increasing switching frequency will be explored in future research.

Fig. 4 compares the performance of SC-TP, SC-SP, and TC-TP topologies over a 10-day continuous operation period during spring, when PV power generation is at its moderate level. Solar power production and load demand are identical across all phases in each scenario, as shown in Fig. 4a. Additionally, Fig. 4b compares each scenario's interaction with the electricity grid. The SC-SP topology exhibits the highest energy exchange and, consequently, the lowest self-sufficiency ratio, whereas the TC-TP topology achieves the lowest energy exchange and the highest self-sufficiency ratio. Notably, phase balance is prioritized as the optimization objective across all scenarios rather than maximizing self-sufficiency, which leads to slightly lower self-sufficiency ratios than if the optimization had focused solely on maximizing self-sufficiency.

Additionally, a comparison of the ES SoC under both single-phase and multi-phase operating conditions is presented in Fig. 4c. The results indicate that when the ER distributes PV-generated power and stored energy across all phases, battery charge and discharge cycles become more frequent, which diminishes ES longevity by increasing stress on the battery cells [25]. Notably, when the ER and ES operate in single-phase mode, the SEMA has fewer opportunities to maximize the usage of locally generated energy and the self-consumption ratio. This leads to an increased need for PV curtailment or grid injection when the ES is fully charged, ultimately resulting in lower ES utilization. However, this approach helps maintain a higher battery health ratio.

A comparison of various operational modes indicates that the TC-TP topology is the least effective at protecting battery cells from rapid degradation, while the SC-SP mode achieves the highest ES health status during operation compared to other modes. Under high solar energy availability, the performance of the SC-TP mode is comparable to, though slightly better than, the TC-TP mode in terms of ES longevity. However, when solar energy is limited, the SC-TP mode outperforms TC-TP by using the ES less frequently.

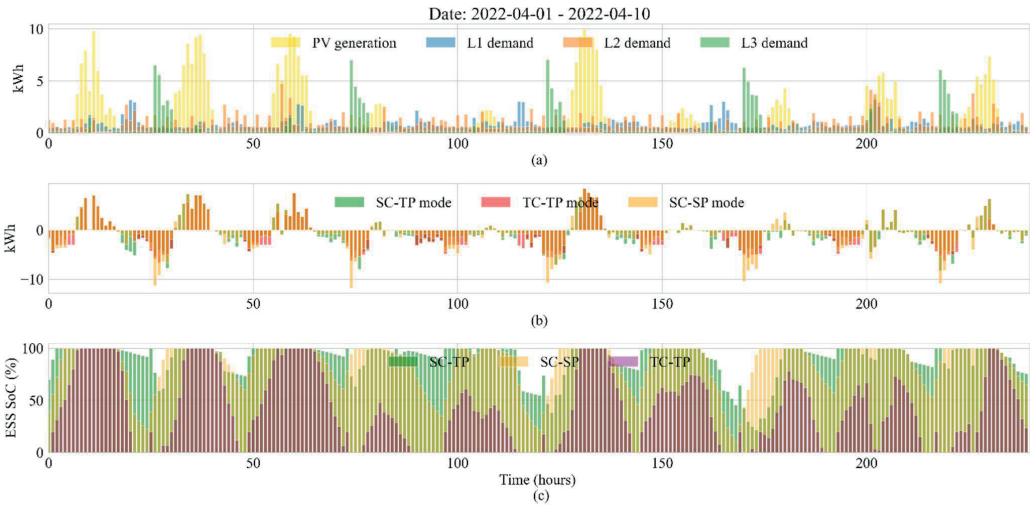


Fig. 4. Comparison of the topologies performances during first 10 days in May-2022. (a) Load demand and PV generation. (b) Amount of energy exchange with the electricity grid (negative values show imported energy and positive values show exported energy to the electricity grid). (c) ES SoC (%) level during systems' operation.

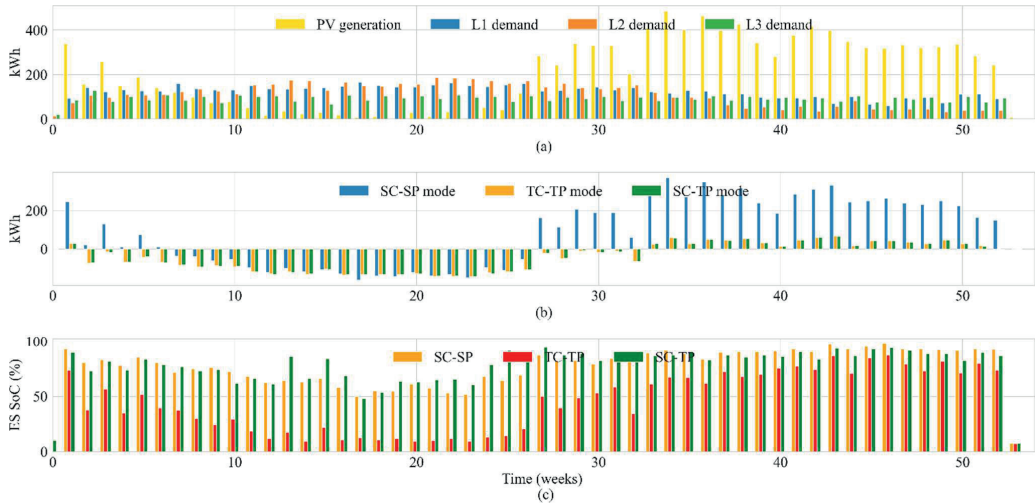


Fig. 5. Comparison of the topologies based on weekly aggregated performance over a year: (a) Total load demand and PV generation for each week; (b) Total weekly energy exchange with the electricity grid; (c) Average weekly SoC (%) ratio of the ES system.

While this improved longevity is beneficial, it may not fully meet end-user expectations, as reduced ES operation increases reliance on the electricity grid to meet demand. Therefore, it is essential for the energy management unit to account for ES degradation costs and prioritize resource allocation based on electricity tariffs.

Fig. 5 presents a comparison of the performance of the investigated operational modes over a year. From a broader perspective, the differences between topologies become clearer. For example, during winter—particularly between weeks 10 and 20—when solar generation is exceptionally low, grid interactions are quite similar across modes. However, during peak solar generation—particularly between weeks 30

and 50—the SC-SP topology shows significantly higher grid interactions, leading to increased grid power flow and potential congestion. This outcome further highlights the limitations of the SC-SP topology in maximizing PV self-consumption, as most of the on-site generated solar power is injected back into the grid. However, two other topologies have been able to utilize locally available renewable sources, to minimize their energy exchanges with the electricity network. The differences between topologies interaction with the electricity network is demonstrated in Fig. 5b.

Fig. 5c compares the average SoC (%) levels across topologies. It is evident that during periods of low solar generation, TC-TP utilizes ES more frequently than the other two topologies. However, when solar generation increases, the SoC (%) levels between topologies become more similar, though TC-TP still charges and discharges the batteries more often. Interestingly, the performance of SC-SP and SC-TP topologies is almost identical, with both benefiting from ES in an equivalent manner. Finally, Fig. 6. shows number of times the ER linked renewable resources to each phase in SC-TP topology.

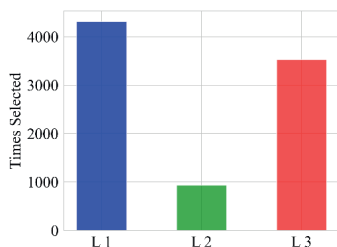


Fig. 6. A number of times, each phase is connected through ER to the local renewable energy sources.

IV. COST BENEFITS ANALYSIS

Fig. 7 presents a comparison of the cost distribution between different ER components in case of SC-TP and TC-TP.

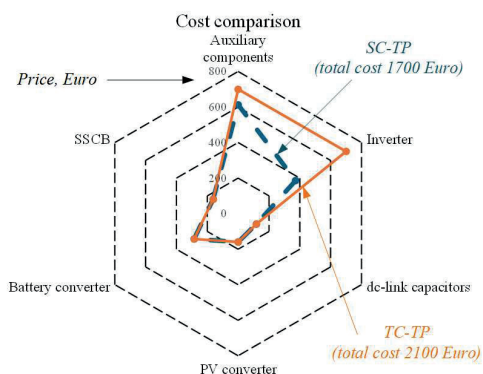


Fig. 7. Cost comparison diagram which shows the cost distribution between different ER components in case of SC-TP and TC-TP.

The cost analysis is based on the retail price of the components used for a single prototype. It includes the cost of all semiconductors and passive components, such as heatsinks and inductors. However, the cost of the printed circuit board and the enclosure used for the demonstrator are not included in the calculations, as these costs are not representative and strongly depend on scaling.

The prototype of SC-TP ER is shown in Fig. 1., and the total cost of the considered components is approximately 1700 euros. The diagram shows that the most expensive part of the prototype relates to the auxiliary (common) components, such as power supply circuits, heatsinks, and relays.

The prototype of the TC-TP ER was not assembled or used in real tests. However, its cost was evaluated based on a bill of materials collected to design the prototype. The total power of the three-phase inverter, as well as the power of the single-phase inverter, was the same, following the same concept with a common ground approach. The overall cost of the components included in the analysis is around 2100 euros, which is 25% higher than that of the SC-TP ER. This cost increase is attributed to the higher number of components required for redundant circuits to connect and link PV and ESS to all three phases.

It should be noted that the absolute values presented in this work cannot be directly used for primary cost estimation, as they strongly depend on scaling, the supply chain during production, and auxiliary circuit optimization. However, the relative comparison is reliable and can be used for cost analysis.

V. CONCLUSIONS

The proposed SC-TP topology improves PV self-consumption and the phase unbalance ratio by 23% and 16% respectively, compared to the SC-SP solution. Moreover, it offers a significant cost saving of 25% compared to the TC-TP system. This achievement underscores the potential benefits of SC-TP systems as a promising solution to replace SC-SP topologies. Since the SC-TP topology offers better phase balancing, it effectively mitigates the negative effects of distributed behind-the-meter PV setups, improves grid reliability and resiliency, and better distributes the load across each phase. This highlights its value in enhancing grid stability and reducing operational challenges for utility providers.

ACKNOWLEDGMENTS

This work was supported by the Estonian research council grants PRG675, EAG234 and PRG1463.

REFERENCES

- [1] I. Calero, C. A. Cañizares, K. Bhattacharya and R. Baldick, "Duck-Curve Mitigation in Power Grids With High Penetration of PV Generation," in *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 314-329, 2022.
- [2] Hokmabad, H.N., Husev, O., Belikov, J., Vinnikov, D. and Petlenkov, E., "Energy Storage and Forecasting Error Impact Analysis In Photovoltaic Equipped Residential Nano-Grids," in *IEEE 17th*

- International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, pp. 1-6, 2023.
- [3] D. Q. Hung and Y. Mishra, "Impacts of single-phase PV injection on voltage quality in 3-phase 4-wire distribution systems," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1–5.
 - [4] E. A. S. Ducoin, Y. Gu, B. Chaudhuri, and T. C. Green, "Analytical Design of Contributions of Grid-Forming and Grid-Following Inverters to Frequency Stability," in *IEEE Transactions on Power Systems*, vol. 39, no. 5, pp. 6345–6358, 2024.
 - [5] K. Ma, R. Li, I. Hernando-Gil and F. Li, "Quantification of Additional Reinforcement Cost From Severe Three-Phase Imbalance," in *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 4143-4144, 2017.
 - [6] A. A. Mohamed, C. Sabillon, A. Golriz, M. Lavorato, M. J. Rider, and B. Venkatesh, "Capacity Market for Distribution System Operator – With Reliability Transactions – Considering Critical Loads and Microgrids," *IEEE Transactions on Power Delivery*, vol. 38, no. 2, pp. 902–916, 2023.
 - [7] S. A. Raza and J. Jiang, "Mathematical Foundations for Balancing Single-Phase Residential Microgrids Connected to a Three-Phase Distribution System," in *IEEE Access*, vol. 10, pp. 5292-5303, 2022.
 - [8] S. Lim and J. -W. Park, "Estimation of Behind-the-Meter Power Loss of Renewables With EMS Data," in *IEEE Transactions on Power Systems*, vol. 38, no. 3, pp. 2974-2977, 2023.
 - [9] F. Bu, R. Cheng, and Z. Wang, "A two-layer approach for estimating behind-the-meter PV generation using smart meter data," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 885–896, 2022.
 - [10] N. Kouvelas and R. V. Prasad, "Efficient Allocation of Harvested Energy at the Edge by Building a Tangible Micro-Grid—The Texas Case," in *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 1, pp. 94-105, 2021.
 - [11] J. Huang, X. Wang, Y. Wang, C. Shao, G. Chen, and P. Wang, "A Game-Theoretic Approach for Electric Vehicle Aggregators Participating in Phase Balancing Considering Network Topology," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 743–756, 2024.
 - [12] S. Liu *et al.*, "Practical Method for Mitigating Three-Phase Unbalance Based on Data-Driven User Phase Identification," in *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1653-1656, 2020.
 - [13] P. Li, W. Wu, X. Wang, and B. Xu, "A Data-Driven Linear Optimal Power Flow Model for Distribution Networks," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 956–959, 2023.
 - [14] D. Younsi, H. E. Z. Farag, H. Zeineldin, and E. El-Saadany, "Optimized Unsymmetrical Per-Phase Droop for Soft Line Switching of Reconfigurable Unbalanced Inverter-Based Islanded Microgrid," *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 3851–3866, 2024.
 - [15] X. Cui, G. Ruan, F. Vallée, J. -F. Toubeau and Y. Wang, "A Two-Level Coordination Strategy for Distribution Network Balancing," in *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 529-544, 2024.
 - [16] M. Lu, W. Cai, S. Dhople, and B. Johnson, "Large-signal stability of phase-balanced equilibria in single-phase grid-forming inverter systems," *IEEE Trans. on Power Electronics*, vol. 39, no. 3, pp. 3623–3636, 2024.
 - [17] L. Fu, W. Wang, Z. Y. Dong and Y. Li, "Optimal Reconfiguration for Active Distribution Networks Incorporating a Phase Demand Balancing Model," *IEEE Transactions on Power Systems*, vol. 39, no. 5, pp. 6183-6195, 2024.
 - [18] J. Huang, X. Wang, Y. Wang, C. Shao, G. Chen and P. Wang, "A Game-Theoretic Approach for Electric Vehicle Aggregators Participating in Phase Balancing Considering Network Topology," in *IEEE Transactions on Smart Grid*, vol. 15, no. 1, pp. 743-756, 2024.
 - [19] L. Fang, K. Ma, F. Li and F. Xue, "Novel Incentive Scheme to Motivate Flexible Customers for Phase Balancing," in *CSEE Journal of Power and Energy Systems*, vol. 8, no. 4, pp. 1048-1059, 2022.
 - [20] Hokmabad HN, Husev O, Kurnitski J, Belikov J. "Optimizing size and economic feasibility assessment of photovoltaic and energy storage setup in residential applications," *Sustainable Energy, Grids and Networks*, 38:101385, 2024.
 - [21] S. Rahimpour, "Common-Ground Energy Router Structure with Enhanced Reliability and Protection," Ph.D. dissertation, *Tallinn University of Technology*, 2024.
 - [22] Shahsavari, T.H., Kurdkandi, N.V., Husev, O., Babaei, E., Sabahi, M., Khoshkbar-Sadigh, A. and Vinnikov, D., "A New Flying Capacitor-Based Buck–Boost Converter for Dual-Purpose Applications." *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, 4(2), pp.447-459, 2023.
 - [23] H. N. Hokmabad, N. Shabir, V. Astapov, E. Petlenkov, O. Husev and J. Belikov, "Feasibility Study of a DC House Connected to a Conventional AC Distribution Network," 2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), Gdynia, Poland, pp. 1-6, 2024.
 - [24] Boscaino, V., Ditta, V., Marsala, G., Panzavecchia, N., Tine, G., Cosentino, V., Cataliotti, A. and Di Cara, D., "Grid-connected photovoltaic inverters: Grid codes, topologies and control techniques." *Renewable and Sustainable Energy Reviews*, 189, p.113903, 2024.
 - [25] M. Amini, M. H. Nazari and S. H. Hosseinian, "Optimal Scheduling and Cost-Benefit Analysis of Lithium-Ion Batteries Based on Battery State of Health," *IEEE Access*, vol. 11, pp. 1359-1371, 2023.

Publication V

H. N. Hokmabad, T. H. Shahsavar, P. P. Vergara, O. Husev, and J. Belikov, 'Forecast-Driven Scenario Generation for Building Energy Management Using Stochastic Optimization', *arXiv [eess.SY]*. 2025.

Forecast-Driven and Scenario-Based Building Energy Management Using a Stochastic Optimization Approach

Hossein Nourollahi Hokmabad, *Student Member, IEEE*, Tala Hemmati Shahsavari, *Member, IEEE*,
Pedro P. Vergara, *Senior Member, IEEE*, Oleksandr Husev, *Senior Member, IEEE*, and Juri Belikov,
Senior Member, IEEE

Abstract—Buildings are essential components of power grids, and their energy performance directly affects overall power system operation. This paper presents a novel stochastic optimization framework for building energy management systems, aiming to enhance buildings' energy performance and facilitate their effective integration into emerging intelligent power grids. In this method, solar power generation and building electricity demand forecasts are combined with historical data, leveraging statistical characteristics to generate probability matrices and corresponding scenarios with associated probabilities. These scenarios are then used to solve the stochastic optimization problem, optimizing building energy flow while accounting for existing uncertainties. The results demonstrate that the proposed methodology effectively manages inherent uncertainties while maintaining performance and outperforming rule-based and custom-built reinforcement learning-based solutions.

I. INTRODUCTION

With the integration of Solar Photovoltaic (PV) systems in buildings and the widespread deployment of Distributed Energy Resources (DERs), local energy management agents are becoming crucial in the energy transition. However, these platforms face challenges in optimizing energy use due to the intermittent nature of renewables and the variability of building electricity demand [1]. The inclusion of Battery Energy Storage (BES) and Electric Vehicles (EVs) further increases system complexity by introducing additional parameters that complicate decision-making [2].

Building energy management methodologies comprise rule-based, deterministic, stochastic, Model Predictive Control (MPC), and Artificial Intelligence (AI)-based approaches, including Reinforcement Learning (RL). Rule-based methods dominate commercial applications due to their simplicity but offer limited adaptability. Deterministic approaches leverage predictions for optimization but struggle with uncertainty and rely on forecast accuracy. Stochastic methods address uncertainties in power flow optimization but require high computational effort due to scenario-based modeling.

Stochastic optimization algorithms include heuristic approaches, such as Monte Carlo methods, which use random sampling to account for uncertainty in optimization problems. These methods generate a wide range of potential solutions

based on probabilistic simulations [3]. Metaheuristic algorithms, such as genetic algorithms [4] and particle swarm optimization [5], are widely adopted approaches for finding near-optimal solutions to complex optimization problems.

Real-time optimization methods are essential for the practical implementation of building EMSs. In this category, stochastic MPC-based systems provide promising results due to their ability to compensate for forecasting errors in real-time. The authors in [6], proposed an MPC-based approach with performance very close to ideal conditions. In [7], a real-time Mixed Integer Linear Programming (MILP)-based stochastic EMS is presented. In [8], the authors utilized stochastic optimization to design a self-healing EMS capable of handling real-time contingencies based on its available resources. Model-free and data-driven solutions are also gaining significant attention for solving building energy optimization problems. With advancements in computing power, AI- and RL-based methods are being widely applied and have demonstrated reliable performance. For instance, RL-based solutions inherently handle system uncertainties. The Deep-RL-based model-free method proposed in [9], outperforms stochastic programming-based methods for energy optimization inside the building. However, the study did not compare the computational complexity of the proposed solution, or the amount of data required for model training. Despite the progress in building EMSs, this field is still relatively new and requires further research to develop reliable solutions that can be confidently integrated into real-world platforms while ensuring performance that justifies the owner's investment. One limitation of state-of-the-art solutions is their tendency to overlook the benefits of incorporating outputs from deterministic forecasting models [10].

To further enhance the performance of stochastic programming solutions, this work fills this gap and presents a forecast-driven approach for generating scenarios for both demand and PV power production to optimize power flow and improve EMS's performance in buildings equipped with PV systems and BES. The proposed approach utilizes historical data distributions to calculate probabilities and pairs generation, and demand profiles based on probability-aware sampling of feasible solutions. The most probable scenarios are then selected and used to solve a multi-scenario optimization problem formulated as a MILP. By incorporating

The work of J. Belikov and H. N. Hokmabad was partially supported by the Estonian Research Council (grant No. PRG1463), and by the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230) funded by the Estonian Ministry of Education and Research.

H. NH. is with Tallinn University of Technology, Tallinn, Estonia (e-mail: hossein.nourollahi@taltech.ee).

T. HS. is with Tabriz University, Tabriz, Iran (e-mail: tala.hemmati@gmail.com).

P. PV is with Delft University of Technology, Delft, Netherlands (email: p.p.vergarabarris@tudelft.nl).

O. H. is with Gdansk University of Technology, Gdansk, Poland (e-mail: oleksandr.husev@pg.edu.pl).

J. B. is with Tallinn University of Technology, Tallinn, Estonia (mail: juri.belikov@taltech.ee).

forecasting tools into the scenario generation and selection procedure, the approach enhances system performance and provides a robust solution that accounts for uncertainties.

II. PROPOSED METHODOLOGY

A. Building Energy Management Systems

Near Zero Energy Buildings' electricity networks generally consist of a Solar PV system, Battery Energy Storage, power electronics infrastructure, and loads. Power electronics tools are essential for enabling power exchange and conversion within the building network and with the electricity grid.

Building EMSs serves as a supervisory element, collecting near real-time indigenous data from the building's electricity network using Internet of Things (IoT) sensors and power electronics devices, as well as the required external data, such as weather conditions and electricity tariffs, from the internet. These data are utilized to optimize energy flow within the building to achieve predefined goals, such as minimizing energy costs or maximizing self-consumption. Figure 1 provides a schematic representation of building EMS and its components.

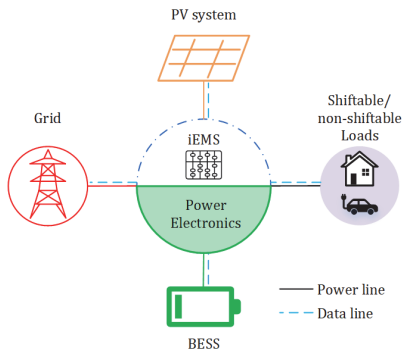


Fig. 1. Schematic representation of near zero energy buildings.

In buildings, loads are usually divided into two groups: shiftable and non-shiftable. Shiftable loads are those that the management system has the flexibility to reschedule based on optimization outcomes. Examples of shiftable loads include washing machines, robotic vacuum cleaners, dishwashers, and heating ventilation systems. Non-shiftable loads, on the other hand, are those that the EMS cannot modify and must be served as requested. Examples of non-shiftable loads include entertainment devices, laptops, and lighting. Additionally, some loads can be categorized as hybrid loads, such as EV chargers and heat pumps, as they can function as either shiftable or non-shiftable loads depending on user requirements. In this paper, we consider all loads to be non-shiftable since they are typically the dominant type of loads in residential applications.

B. Demand and Renewable Energy Generation Forecasts

Any optimal decision-making process or algorithm requires insight into the system's future input variables and states. Without accurate forecasting, making optimal decisions becomes highly unlikely. Conversely, if perfect forecasts were available, stochastic optimization problems could be simplified into a deterministic format. However, the intermittent and stochastic nature of electricity demand and

Solar PV power generation presents significant challenges to prediction accuracy.

Since discussions on forecasting methods and models are beyond the scope of this work, we have used the day-ahead solar PV power generation forecasting models proposed in [11] and Long Short-Term Memory (LSTM) based Deep Neural Networks (DNN) model for demand forecasting. These models generate 24 values per run, each with a 1-hour resolution for the next 24 hours. The generated values represent the expected mean values for solar PV power generation and building electricity demand at each hour of the day.

C. Forecast-Driven Scenario Generation

As mentioned above, forecasting tools generate 24 predictions for electricity generation and consumption for the next day. Let \bar{F}_h^G and \bar{F}_h^D represent the forecasted values for power generation and demand at hour h , respectively. Assuming the forecasting models provide acceptable accuracy, these values should serve as the best possible estimates of the system's uncertain input parameters for the next 24 hours. However, the inherent uncertainties must be addressed to ensure robust decision-making.

To this end, based on available historical records and data distribution the mean and standard deviation (σ) are calculated for each hour of the day. After obtaining these values, the initial mean is replaced with the forecasted values (\bar{F}_h^G and \bar{F}_h^D). Assuming a Gaussian distribution, for each hour of the day, a normal distribution curve is generated using the updated mean and the previously computed σ . Since these curves represent physical quantities with finite values, they are constrained within the minimum and maximum possible ranges for electricity generation and demand.

Then, the covered range is divided into $\mathcal{R} = 100$ sections, and for each section, the probability of the actual measured value falling within that range is calculated. For instance, assuming the maximum power generation capacity of the solar PV system is 5 kWp, each section will have a resolution of 50 W. By limiting the number of sections to a fixed value, regardless of sizes of PV systems and building demand, the computational complexity remains consistent across all cases. By dividing the continuous range of possible values, we discretize and limit the potential subsequent values. However, the impact of this action is negligible in system performance.

After calculating these probabilities, a matrix of forecasting probabilities is constructed. Let G_p^F and D_p^F represent the matrices for the probabilities of potential values for electricity generation and demand, respectively, with dimensions $\mathcal{R} \times \mathcal{D}$, where \mathcal{D} is the forecasting horizon, which in this case equals 24. Similarly, a probability matrix is generated for each hour of the day using previously recorded data on solar PV power generation and demand. Let G_p^H and D_p^H represent the matrices for the probabilities of historical values for generations and demand, respectively, with dimensions $\mathcal{R} \times \mathcal{D}$.

The calculation of the D_p^H matrix is straightforward. First, recorded demand values are clustered based on their respective hours of the day and then grouped according to their corresponding power range. Once all records are classified, the probabilities for each hour and power range are computed.

However, since solar PV power generation is highly dependent on weather conditions, classifying records based solely on temporal data would lead to inaccurate results, as solar irradiance in summer is not comparable to that in winter. To address this, we propose a novel method for classifying solar PV power generation that eliminates seasonal impacts.

To achieve this, for each day of the year and each hour of the day, the maximum possible solar irradiance values are calculated based on the sun position in the sky, and building's latitude, and longitude. These values represent the theoretical maximum under clear sky conditions. The obtained solar irradiance value is then fed into a physics-based simulation of the building's solar PV system to determine the potential maximum power generation, the detailed information about physics-based modeling can be found in [11]. The minimum possible solar PV power generation value is derived from historical data by identifying the lowest recorded value for the same day and hour within a ± 30 -day window.

Using the obtained minimum and maximum ranges for each hour and day of the year, recorded Solar PV power generation values can be categorized into a predefined number of classes. This is done by normalizing the range, determining class boundaries, and assigning each measurement to its corresponding class. In this approach, measurements are classified based on the percentage of Solar PV power generation relative to the maximum feasible value. This eliminates the seasonality factor from the data, allowing for a direct comparison of Solar PV power generation probabilities between summer and winter without considering the absolute magnitude of the data. Figure 2 illustrates the process described.

After assigning all measurements to their corresponding classes, the probabilities for each hour of the day are calculated for the entire dataset based on the classes that share the same hour label. As a result, the matrix $\mathcal{G}_p^{\mathcal{H}}$ will have dimensions equal to the number of hours in a day and the number of classes, which are set to 100 in this case. Then, the final probability matrices are constructed as:

$$\mathcal{G}_p^{\mathcal{T}} + \mathcal{G}_p^{\mathcal{H}} = \mathcal{G}_p^{\mathcal{T}}, \quad \mathcal{G}_p^{\mathcal{T}} \in \mathbb{R}^{\mathcal{R} \times \mathcal{D}}, \quad (1)$$

$$\mathcal{D}_p^{\mathcal{T}} + \mathcal{D}_p^{\mathcal{H}} = \mathcal{D}_p^{\mathcal{T}}, \quad \mathcal{D}_p^{\mathcal{T}} \in \mathbb{R}^{\mathcal{R} \times \mathcal{D}}, \quad (2)$$

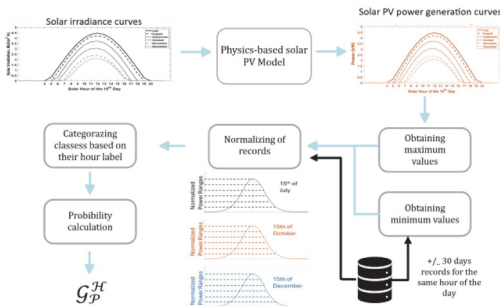


Fig. 2. Abstract representation of obtaining $\mathcal{G}_p^{\mathcal{H}}$ from historical solar PV power generation records. Solar irradiance curves are extracted from estimation of the hourly global solar irradiation based on numerical weather predictions.

where $\mathcal{G}_p^{\mathcal{T}}$ and $\mathcal{D}_p^{\mathcal{T}}$, are total probabilities for production and demand ranges, respectively, considering both historical and forecasted values. Whenever probabilities are accumulated, an averaging operation is also performed to ensure that the total probability sum always remains equal to one.

Scenarios are generated based on combinations of paired power generation and consumption values. To achieve this, the Cumulative Distribution Function (CDF) for each hour of the day is derived from the $\mathcal{G}_p^{\mathcal{T}}$ and $\mathcal{D}_p^{\mathcal{T}}$ probability matrices. For each CDF curve, the probability range is equally divided into \mathcal{S} sections, where \mathcal{S} represents number of scenarios. Then, one random value is generated for each section, and based on these values, the corresponding points are selected for each time step in the control horizon. Assuming $\mathcal{S} = 100$ for each hour of the day, there are 100 power generation ($G_h^{\mathcal{S}}$) and demand ($D_h^{\mathcal{S}}$) values, where $h \in [0, 23]$ and $\mathcal{S} \in [1, \mathcal{S}]$. Finally, the daily consumption and generation profiles are generated by combining $G_h^{\mathcal{S}}$ and $D_h^{\mathcal{S}}$ as:

$$\mathfrak{S} = \begin{bmatrix} (G_0^{\mathcal{S}}, D_0^{\mathcal{S}}) & \cdots & (G_0^{\mathcal{S}}, D_0^{\mathcal{S}}) \\ \vdots & \ddots & \vdots \\ (G_{h-1}^{\mathcal{S}}, D_{h-1}^{\mathcal{S}}) & \cdots & (G_{h-1}^{\mathcal{S}}, D_{h-1}^{\mathcal{S}}) \end{bmatrix}_{h \times \mathcal{S}}, \quad (3)$$

where \mathfrak{S} is the matrix of all generated scenarios. Figure 3 illustrates the described procedure.

Also, since combined generation and consumption values ($G_{h-1}^{\mathcal{S}}, D_{h-1}^{\mathcal{S}}$) have different probabilities, the probability for each pair is considered as multiplication of each individual probability. For calculating the total probability of each scenario ($\bar{P}_s^{\mathcal{T}}$), the probability of each individual hour is accumulated and then averaged. These probabilities are stored in a scenario's probability matrix:

$$P_s^{\mathcal{T}} = [\bar{P}_1^{\mathcal{T}}, \bar{P}_2^{\mathcal{T}}, \bar{P}_3^{\mathcal{T}}, \dots, \bar{P}_s^{\mathcal{T}}], \quad P_s^{\mathcal{T}} \in \mathbb{R}^{1 \times \mathcal{S}}. \quad (4)$$

III. OPTIMIZATION FRAMEWORK

In the previous section, scenarios and their associated probabilities have been generated. In this section, the optimization framework will be described, to optimize building energy flow based on defined objectives.

A. Optimization Function Formulation

The main optimization problem consists of sub-optimization problems for each scenario. In other words, the optimal solution is the one that minimizes the defined cost function while considering all scenarios. However, this does not guarantee that the solution is optimal for each individual scenario. A general optimization problem is defined as:

$$\begin{aligned} \min_x \quad & \sum_{i=1}^{\mathcal{S}} \mathcal{F}_i^{\mathcal{T}} \cdot X_i \cdot \hat{P}_i^{\mathcal{T}} \\ \text{s.t.} \quad & X_{i,\min} \leq X_i \leq X_{i,\max} \end{aligned} \quad (5)$$

where \mathcal{F}_i is the vector of optimization variables, X_i represents the vector of optimization factors, and $\hat{P}_i^{\mathcal{T}}$ denotes the probability of each scenario's occurring within the corresponding optimization horizon. Additionally, \mathcal{S} represents the number of selected candidate scenarios. To manage the complexity of the optimization process, only the 10 most probable scenarios are selected from the generated set.

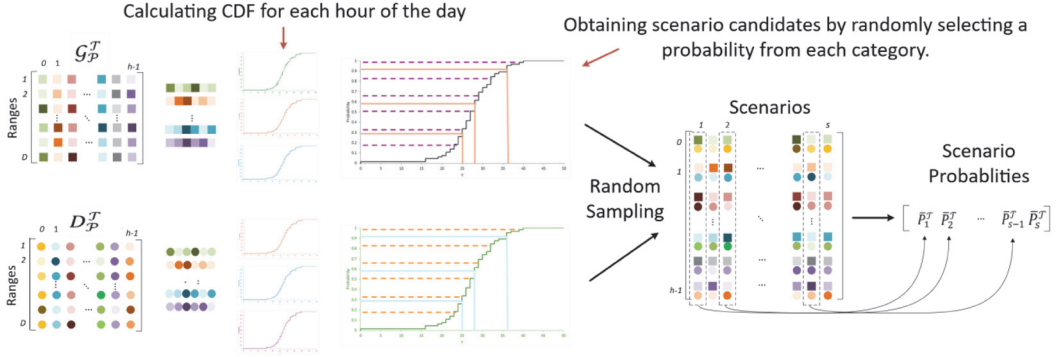


Fig. 3. Daily scenario generation is performed based on calculated probability matrices and random candidate selection. For each hour of the day, probabilities for $d=100$, different possible values are accumulated to form CDF curves. From each range, candidate values are selected through random sampling and combined to form complete daily profiles. Finally, the 10 most probable scenarios are selected for the final optimization round.

after sorting them by probability. It is also worth noting that the probabilities of the selected scenarios are normalized to reflect their relative differences. Furthermore, the sub-optimization problem for each scenario is defined as:

$$\min \sum_{h=0}^w E_h^{imp} \times ToU_h^{imp} - E_h^{exp} \times ToU_h^{exp}. \quad (6)$$

Since energy is the time integral of power, under the assumption of constant time intervals and stable system voltage and current levels within each interval, energy parameters can be expressed as $E = P \times t$. Consequently, the optimization function can be reformulated and solved based on the power flow within the building's internal electricity network. Thus, the objectives and constraints for Eq. (6) can be represented as:

$$\forall h, i, 0 \leq P_{A \rightarrow B}^i[h], \quad (7)$$

$$\forall h, i, 0 \leq P_{pv \rightarrow gr}^i[h] + P_{pv \rightarrow es}^i[h] + P_{pv \rightarrow ld}^i[h] \leq P_{pv, max}, \quad (8)$$

$$\forall h, i, 0 \leq P_{gr \rightarrow ld}^i[h] + P_{gr \rightarrow es}^i[h] \leq P_{gr, max}, \quad (9)$$

$$\forall h, i, 0 \leq P_{es \rightarrow ld}^i[h] + P_{es \rightarrow gr}^i[h] \leq P_{es, max}, \quad (10)$$

$$\forall h, i, P_{pv \rightarrow ld}^i[h] + P_{pv \rightarrow gr}^i[h] + P_{pv \rightarrow es}^i[h] = P_{pv}^i[h], \quad (11)$$

$$\forall h, i, P_{es \rightarrow ld}^i[h] + P_{gr \rightarrow ld}^i[h] + P_{pv \rightarrow ld}^i[h] = P_{ld}^i[h], \quad (12)$$

$$\forall h, i, SoC_{min} \leq SoC_{es}^i[h] \leq SoC_{max}, \quad (13)$$

$$\forall h, i, export_{gr}^i[h] + import_{gr}^i[h] \leq 1, \quad (14)$$

$$\forall h, i, charge_{es}^i[h] + discharge_{es}^i[h] \leq 1, \quad (15)$$

where $E_h^{imp} = (P_{gr \rightarrow ld}^i[h] + P_{gr \rightarrow es}^i[h]) \times t$ is the total imported energy for each hour of the day from the grid, and $E_h^{exp} = (P_{pv \rightarrow gr}^i[h] + P_{es \rightarrow gr}^i[h]) \times t$ represents the total amount of net energy exported to the grid during hour (h). Also, ToU_h^{imp} , and ToU_h^{exp} denote the time-of-use tariffs for imported and exported energy, respectively. The notation $P_{A \rightarrow B}^i[h]$ represents power flow from point A to point B, where $A, B \in \{pv, gr, es\}$. Equation (7) ensures that all power flows are non-negative. Furthermore, $P_{pv \rightarrow gr}^i[h]$, $P_{pv \rightarrow es}^i[h]$, and $P_{pv \rightarrow ld}^i[h]$ represent power flow from PV to grid, energy storage, and loads, respectively. The notation $P_{pv, max}$ denotes

the maximum allowable power output from the PV system, constrained by the PV system size and power electronics limitations. $P_{gr \rightarrow ld}^i[h]$, and $P_{gr \rightarrow es}^i[h]$, represent power flow from grid to loads and energy storage, respectively, and $P_{gr, max}$ is the maximum allowable power exchange with the grid. $P_{es \rightarrow ld}^i[h]$, and $P_{es \rightarrow gr}^i[h]$ represent power flow from energy storage to load and grid, respectively, and $P_{es, max}$ is the maximum charge/discharge power of the energy storage unit.

Equations (11) and (12) are equality constraints ensuring that the optimization algorithm satisfies demand and utilizes all available solar PV power under all conditions. Here, $P_{pv}^i[h]$ and $P_{ld}^i[h]$ represent the generated and demanded power, at time h , respectively. In Eq (13), SoC_{min} , SoC_{max} define, respectively, the minimum and maximum allowable battery State of Charge (SoC) levels. Finally, equations (14) and (15) prevent the optimization algorithm from generating infeasible solutions. For example, importing and exporting power to the grid simultaneously is physically impossible. Therefore, $export_{gr}^i[h]$, and $import_{gr}^i[h]$ are Boolean values, that enforce this constraint. A similar logic applies to energy storage, where $charge_{es}^i[h]$, and $discharge_{es}^i[h]$ are Boolean variables indicating the charging or discharging state of the battery. If the battery is charging $charge_{es}^i[h]=1$, and otherwise the $discharge_{es}^i[h]=1$.

C. Performance Metrics

Various factors can be considered when evaluating the performance of EMSs. Depending on optimization goals and problem formulation, these factors may include self-consumption ratio, electricity costs, energy storage utilization and charge/discharge cycles, demand response efficiency, energy conversion losses, and more. In this study, we focus on two key performance metrics: electricity costs as the primary performance indicator and building self-consumption ratio ($\delta \in [0, 100]$ %) as the secondary factor. However, since the optimization problem is formulated solely based on minimizing the energy bill, δ serves only as a performance measurement metric and does not influence the optimization process.

Annual energy bill (\mathcal{A}_{bill}) is defined as:

$$A_{bill} = \sum_{d=1}^{365} \sum_{h=0}^w E_{h,d}^{imp} \times ToU_{h,d}^{imp} - E_{h,d}^{exp} \times ToU_{h,d}^{exp}. \quad (16)$$

And an annual self-sufficiency ratio is defined as:

$$\sigma = \frac{\sum_{d=1}^{365} \sum_{h=0}^{23} \frac{E_{pv \rightarrow ld}[h,d] + E_{pv \rightarrow es \rightarrow ld}[h,d]}{E_{pv}[h,d]}, \quad (17)$$

where $E_{pv \rightarrow ld} = P_{pv \rightarrow ld} \times t$, $E_{pv} = P_{pv} \times t$, and $E_{pv \rightarrow es \rightarrow ld} = P_{pv \rightarrow es \rightarrow ld} \times t$. The notation $E_{pv \rightarrow es \rightarrow ld}$ represents the amount of energy generated by the solar PV system, stored in the energy storage system, and later delivered to the load. This term is often overlooked in literature, where the self-consumption ratio is typically calculated only by considering the real-time power delivery from the PV system to the load.

D. Performance Benchmarks

The proposed algorithm is benchmarked against three different approaches. The first approach is a simple rule-based algorithm that follows a priority-based energy management strategy. It prioritizes supplying demand from the PV system first, then from the BES, and finally from the grid. If the generated power exceeds the demand, the surplus energy is stored in the battery. If the battery's SoC reaches its maximum limit, any additional energy is exported to the grid. Conversely, during energy shortages, the system first utilizes stored energy in the battery, and if the demand is still not met, the remaining energy is imported from the grid.

The second approach assumes that the optimization algorithm has access to ideal forecasts for PV power generation and demand over the next 24 hours. This scenario represents the best feasible solution for an optimization problem, as it eliminates performance losses due to uncertainty, given that all future information is known and predictable. Finally, two RL-based agents, one using Proximal Policy Optimization (PPO) and the other based on Deep Q-Networks (DQN), are utilized for benchmarking.

In all scenarios, the system follows a consistent strategy. At the start of the day, optimal control signals are generated based on forecasts for PV power generation and demand. Throughout the day, a high-resolution algorithm ensures that voltage and current levels remain within acceptable ranges. If the system falls short (e.g., PV power generation is insufficient), the grid is used to compensate for energy shortages, ensuring continuous operation within safe parameters. And if the PV generation exceeds the demand, the extra generation will be first directed to charge the BES and then the main grid if the BES SoC level reaches SoC_{max} .

IV. NUMERIC RESULTS

A. Case Study

Four years of historical consumption data from House Number 1, obtained from the UK-DALE dataset [12] and recorded from March 2013 to April 2017, are combined with synthetic Solar PV generation data for a similar period and the same geographical location, generated using the Photovoltaic Geographical Information System (PVGIS) [13]. The generated dataset's granularity is 1 hour, and the recorded values represent the average measurement during each hour of the day. Additionally, Time-of-Use (ToU) electricity price

data, sourced from Estonia's end-user electricity price dataset and recorded from March 2020 to April 2024, are utilized.

The optimization problem is formulated as a MILP model. The problem has been implemented using the Pyomo optimization framework and solved using the IPOPT solver. Additionally, all code has been developed using Python 3.11. Table I collects systems parameters, and constraints.

TABLE I. BUILDING ELECTRICITY NETWORK CHARACTERISTICS AND OPERATIONAL CONSTRAINTS

Variable Name	Value	Unit	Symbol
PV system	10	kWp	-
PV inverter size	12	kW	$P_{pv,max}$
BES capacity	10	kWh	-
BES inverter size	5	kW	$P_{es,max}$
Grid-connected converter size	5	kW	$P_{gr,max}$
Minimum SoC	15	%	SoC_{min}
Maximum SoC	90	%	SoC_{max}

B. Probability Matrices

Figure 4 (a) illustrates the number of members for distinct class labels, each corresponding to a specific range that covers the entire span of possible solar PV power generation and demand values. As shown in the figure, the probability of higher solar PV power generation is notably higher during midday hours, which is consistent across different seasons. This reflects the natural behavior of solar power generation, with peak production occurring when the sun is at its highest in the sky.

Regarding the demand patterns, which are depicted in Figure 4 (b), the probability distribution is spread across the day, but with notable peaks during the early morning hours and evening times. These peaks align with typical electricity demand patterns, as buildings typically require more power in the morning for activities such as heating, cooling, and appliance use, and again in the evening when residents return home and start using electricity for lighting, cooking, and other household tasks.

C. Generated Scenarios

Scenarios are generated by randomly selecting and bounding consumption and generation power values for each hour of the day based on initial day-ahead forecasts. Figure 5 shows the predicted PV power generation and demand for 72 hours during the system's operation in the first week of June 2015. Observe that the PV power generation prediction outperforms the power consumption predictions. This can be attributed to the higher randomness and uncertainty associated with building power demand, whereas PV power generation is highly correlated with weather conditions and predictions. Numerical weather predictions are easily accessible from weather service providers, leading to more reliable PV power generation forecasts. Figure 6 demonstrates the ten most probable generated scenarios for the corresponding days during the first week of June 2015. These scenarios are extracted from \mathfrak{S} , which is a paired combination of the G_P^J and

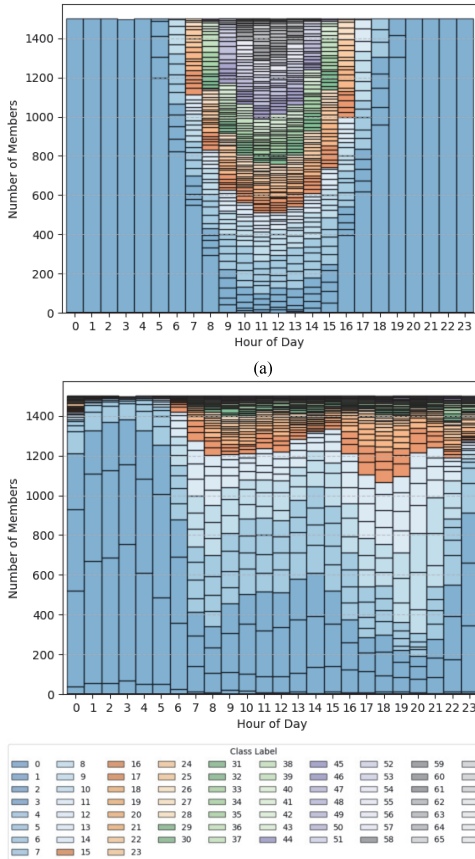


Fig. 4. (a) Solar PV system power generation class membership per hour for entire dataset. (b) Building energy demand class membership per hour for entire dataset. Each graph also represents the distribution of probabilities for each class. The classes with a higher footprint have a higher probability of occurring.

$D_{\mathcal{P}}^{\mathcal{F}}$ probability matrices. As shown, tuning the forecasts with historical data distribution and generating scenarios based on both historical data and forecasts allowed the tool to estimate the range of PV power generation and building demand with moderate accuracy. However, improvements are still needed to ensure that the generated scenarios closely follow real profiles. To ensure realistic forecasting, zero values are assigned to night hours in the final PV production profiles, as power generation during these hours is negligible regardless of forecast outputs or scenario results. This assignment is based on the local sunrise and sunset times.

D. Performance Benchmarking

Each paired scenario is included in Eq. (5), with their corresponding probabilities as a weighting factor. In this work, we have defined the optimization problem to solely minimize the annual electricity bill; however, other factors could be incorporated to construct a multi-objective optimization problem. Figure 7 shows the optimized power flow based on

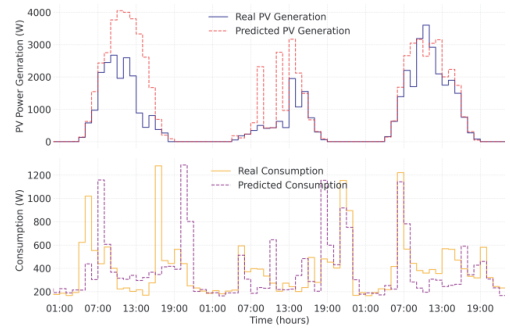


Fig. 5. Real solar PV power generation and building power consumption during the first week of June-2015, in comparison with their day-ahead prediction.

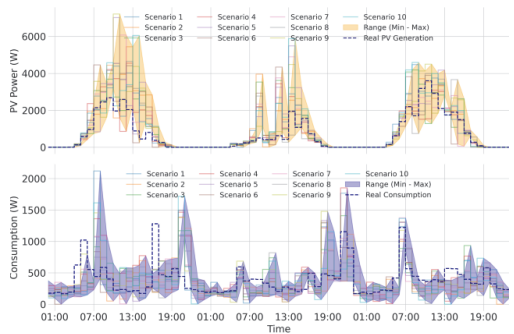


Fig. 6. Ten most probable scenarios based on day-ahead forecasts and historical data distribution.

input variables and scenarios. The decision made for bill minimization is highly dependent on the infrastructure size. Since the focus here is on the algorithm's performance, the results are reported considering only one configuration.

It is noticeable that the proposed methodology makes optimization decisions that are closest to the ideal forecasting scenario compared to other methodologies. However, perfect alignment is not achievable, as it is impossible to precisely forecast the system's future states. Table II presents economic and technical metrics based on one year of system operation. As expected, the ideal forecasting scenario yields the best performance both technically and economically. The closest performance is achieved by the proposed solution, which not only generates revenue but also effectively utilizes the BES to enhance the self-sufficiency ratio.

The proposed solution and the RL-DQN method exhibit relatively similar performance. However, their operational principles and computational complexities differ significantly. The primary computational burden of RL-based approaches lies in the training phase rather than inference. In contrast, the proposed method requires the calculation of probabilities and corresponding scenarios at least once per day, followed by solving an optimization problem that considers multiple scenarios. This computational demand may be regarded as a limitation of stochastic programming-based approaches. A similarity between the two approaches is their reliance on

substantial historical data. Both methods face challenges related to cold-start issues, as their performance heavily depends on the availability of sufficient past data for training or scenario generation. To compare the computational complexity of the proposed method with its benchmarks, all algorithms were executed on the same Windows machine with the following configuration: Intel Core Ultra 9-185H 2.5 GHz processor and 32 GB of RAM. Table III compares the execution times. The reported inference time refers to the complete processing of the test set.

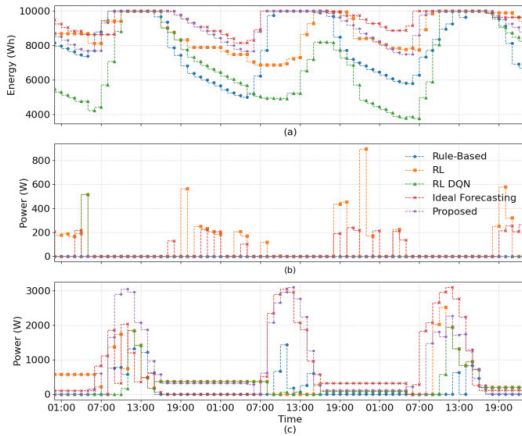


Fig. 7. The optimized power flow comparison between various methods. (a) Hourly-averaged BES stored energy levels, (b) hourly-averaged imported power from the grid, and (c) averaged exported power to the grid.

TABLE II. PERFORMANCE BENCHMARKS

Optimization Method	SCR* (%)	AEB* (€)	ABC* (%)	TIEG* (kWh)	TEEG* (kWh)
Rule-based	82.01	135.67	49.38	2990.50	5451.72
RL-based (PPO)	67.02	5.17	65.64	5615.76	6426.80
RL-based (DQN)	81.48	-70.73	48.94	3078.95	9384.62
Ideal forecasting	91.44	-193.85	71.26	746.82	3224.71
Proposed	81.10	-97.48	44.08	2721.38	4683.14

*SCR: PV Self-Consumption Ratio, AEB: Annual Electricity Bill, ABC: Average Battery Charge, TIEG: Total Imported Energy from the Grid, TEEG: Total Exported Energy to the Grid.

V. CONCLUSION

This article presents a forecast-driven stochastic energy management agent for building EMSs. The proposed solution relies on day-ahead forecasts and historical data distributions to generate the most probable scenarios for solar PV power generation and building electricity demand. These scenarios are then used for power flow optimization to minimize the annual electricity bill while accounting for inherent system uncertainties. The performance of the proposed method is compared with rule-based, two RL-based methods, and ideal forecasting approaches. The results confirm the superiority of the proposed method over the benchmarked models. However, it should be noted that the performance of RL-based methods heavily depends on the learning process and

hyperparameters. The main disadvantages of the proposed solution are its relatively complex computational process and the need for extensive historical data, which is a typical challenge for stochastic programming optimization-based methods. In future work, cold start issues and a sensitivity analysis of the amount of data required to achieve satisfactory performance will be investigated.

TABLE III. EXECUTION TIME COMPARISON

Optimization Method	Train (ms)	Inference (ms)
Rule-based	-	350
RL-based (PPO)	38197	7710
RL-based (DQN)	59959	6457
Ideal forecasting	-	18974
Proposed	-	62531

REFERENCES

- [1] H. N. Hokmabad, O. Husev, J. Kurnitski, and J. Belikov, 'Optimizing Size and Economic Feasibility Assessment of Photovoltaic and Energy Storage Setup in Residential Applications', *Sustainable Energy, Grids and Networks*, vol. 38, p. 101385, 2024.
- [2] A. Raza, L. Jingzhao, Y. Ghadi, M. Adnan, and M. Ali, 'Smart Home Energy Management systems: Research Challenges and Survey', *Alexandria Engineering Journal*, vol. 92, pp. 117–170, 2024.
- [3] I. Mezghani, S. Misra, and D. Deka, 'Stochastic AC Optimal Power Flow: A Data-driven Approach', *Electric Power Systems Research*, vol. 189, p. 106567, 2020.
- [4] R. El Makroum, A. Khallaayoun, R. Lghoul, K. Mehta, and W. Zörner, 'Home Energy Management System based on Genetic Algorithm for Load Scheduling: A Case Study based on Real Life Consumption Data', *Energies*, vol. 16, no. 6, p. 2698, 2023.
- [5] C. Menos-Aikateriniadis, I. Lamprinos, and P. S. Georgilakis, 'Particle Swarm Optimization in Residential Demand-side Management: A Review on Scheduling and Control Algorithms for Demand Response Provision', *Energies*, vol. 15, no. 6, p. 2211, 2022.
- [6] M. Yousefi, A. Hajizadeh, M. N. Soltani, and B. Hredzak, 'Predictive Home Energy Management System with Photovoltaic Array, Heat Pump, and Plug-in Electric Vehicle', *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 430–440, 2021.
- [7] A. Chatterjee, S. Paul, and B. Ganguly, 'Multi-objective Energy Management of a Smart Home in Real Time Environment', *IEEE Transactions on Industry Applications*, vol. 59, no. 1, pp. 138–147, 2023.
- [8] A. Akbari-Dibavar, B. Mohammadi-Ivatloo, K. Zare, and A. Anvari-Moghaddam, 'Optimal Scheduling of a Self-healing Building Using Hybrid Stochastic-robust Optimization Approach', *IEEE Transactions on Industry Applications*, vol. 58, no. 3, pp. 3217–3226, 2022.
- [9] J. Suchithra, D. A. Robinson, and A. Rajabi, 'A Model-Free Deep Reinforcement Learning-Based Approach for Assessment of Real-time PV Hosting Capacity', *Energies*, vol. 17, no. 9, p. 2075, 2024.
- [10] A. Stratigakos, S. Camal, A. Michiorri, and G. Kariniotakis, 'Prescriptive Trees for Integrated Forecasting and Optimization Applied in Trading of Renewable Energy', *IEEE Transactions on Power Systems*, vol. 37, no. 6, pp. 4696–4708, 2022.
- [11] H. N. Hokmabad, O. Husev, and J. Belikov, 'Day-ahead Solar Power Forecasting Using LightGBM and Self-attention based Encoder-Decoder Networks', *IEEE Transactions on Sustainable Energy*, pp. 1–13, 2024.
- [12] J. Kelly, "UK Domestic Appliance Level Electricity (UK-DALE) - Disaggregated appliance/aggregated house power [Data set]," Imperial College, 2012. [Online]. Available: <https://doi.org/10.5286/ukerc.edc.000001>.
- [13] Photovoltaic Geographical Information System (PVGIS) EU Science Hub. Available at: https://joint-research-centre.ec.europa.eu/photovoltaic-geographical-information-system-pvgis_en (Accessed: 02 November 2024).

Publication VI

H. N. Hokmabad, N. Shabir, V. Astapov, E. Petlenkov, O. Husev, and J. Belikov, 'Feasibility Study of a DC House Connected to a Conventional AC Distribution Network', in 2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), 2024, pp. 1–6.

Feasibility Study of a DC House Connected to a Conventional AC Distribution Network

Hossein Nourollahi Hokmabad
 Department of Electrical Power
 Engineering and Mechatronics
 Tallinn University of Technology
 Tallinn, Estonia
 honour@taltech.ee

Eduard Petlenkov
 Department of Computer Systems
 Tallinn University of Technology
 Tallinn, Estonia
 eduard.petlenkov@taltech.ee

Noman Shabir
 FinEST Centre for Smart Cities
 Tallinn University of Technology
 Tallinn, Estonia
 noman.shabir@taltech.ee

Oleksandr Husev
 Department of Electrical Power
 Engineering and Mechatronics
 Tallinn University of Technology
 Tallinn, Estonia
 oleksandr.husev@taltech.ee

Victor Astapov
 Department of Electrical Power
 Engineering and Mechatronics
 Tallinn University of Technology
 Tallinn, Estonia
 victor.astapov@taltech.ee

Juri Belikov
 Department of Software Science
 Tallinn University of Technology
 Tallinn, Estonia
 juri.belikov@taltech.ee

Abstract— Leveraging advancements in power electronics, the adoption of Direct Current (dc) technology in net-Zero Energy Buildings (nZEBs) is seen as a promising approach to boost energy efficiency. Emerging dc technology aims to reduce power losses by eliminating unnecessary conversions between dc and Alternating Current (ac). This paper thoroughly assesses the effectiveness of dc and hybrid dc (partial dc) topologies in comparison to conventional ac nZEBs. The hybrid solution integrates ac and dc networks and loads within the building, while the dc solution benefits from a purely dc internal electricity distribution network and exclusively dc loads. The study involves analyzing annual load profile data from 16 neighboring houses in Estonia, simulating their transition to nZEBs with dc, hybrid dc, and pure dc topologies. The study examines power conversion losses, the operational periods of power converters in relation to their maximum power capacity, and energy exchanges with the utility grid. Additionally, it explores the potential for creating energy communities based on the consumption patterns of these houses. The results indicate that pure dc nZEBs might not be as efficient as initially thought, especially when renewable resources are limited.

Keywords—Net zero energy buildings, dc house, ac distribution network, ac-dc hybrid house, renewable energies.

I. INTRODUCTION

Europe's commitment to carbon neutrality by 2050 is set to enhance the integration of both centralized and distributed renewable energy resources. Notably, 95% of renewable energy harvesters in buildings are dc-operating Photo Voltaic (PVs) panels [1]. Likewise, Electric Vehicles (EVs), Energy Storage Systems (ESSs), and electronic appliances primarily use dc technology. Advancements in power electronics and semiconductor technologies are increasingly making electric devices operate in dc mode for higher efficiency. Most modern home and office appliances, being internally dc-based, have adapted to existing ac networks by losing some efficiency. Furthermore, in transmission side, high voltage dc technology is a more efficient solution for long-range power transmission than traditional transformers-based methods.

These trends are reinforcing the move towards dc in energy systems. However, replacing ac-reliant distribution networks, developed over the past 200 years, is not feasible in the short term. Consequently, establishing a pure dc electricity grid is not a realistic option soon. The most anticipated developments are likely to include transitioning conventional grids to smart grids and upgrading buildings with PVs, ESSs, and Energy Management Systems (EMSs). Additionally, installing dc distribution lines alongside existing ac systems

could boost grid capacity. Finally, the creation of 'energy neighbors' would enable the exchange of energy among nearby prosumers and help reduce the load on the grid [2].

Achieving nZEB status requires aligning a building's annual energy consumption and production, necessitating minimized energy use. This can be achieved through advanced construction technologies and cutting unnecessary energy usage. For instance, eliminating unnecessary dc-ac-dc energy conversions potentially improves efficiency by up to 15% [3]. Fig. 1.a, 1.b, and 1.c illustrate conventional ac-based, hybrid ac-dc, and pure dc-based nZEB solutions, respectively.

While the transition to dc technology in buildings compared to other topologies seems to be more efficient, further research is needed to assess the actual performance of dc buildings linked to the existing ac distribution networks. In [4], the required technical standards, and prerequisites for such integration have been collected and reviewed. Work [1], have listed available dc home appliances in the market. The compatibility of existing home appliances with potential dc distribution network, has been evaluated in [5]. The authors recommend 311 V and 230 V as suitable voltage levels within buildings for a dual line dc distribution system. These voltage levels are deemed appropriate for efficiently powering modern devices equipped with switched-mode power supplies

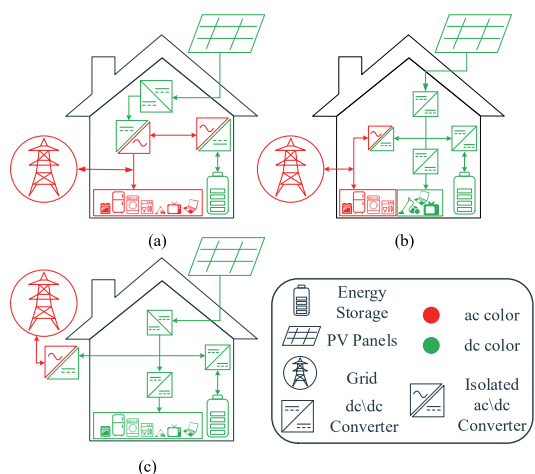


Fig. 1. a) nZEB topology with conventional ac network, b) nZEB topology with hybrid ac-dc network (partial dc), c) nZEB topology with pure dc network. The arrows indicate the direction of power flow.

and universal motors, without the need for significant modifications.

In [6], the efficiency of dc buildings linked to conventional ac distribution networks is explored. The study examines three types of load profiles, representing low, medium, and high-power consumption in US houses. It concludes that higher voltage dc distribution systems are less efficient than lower voltage networks, primarily due to increased losses in ac-dc converters at the 220 V dc range compared to 48 V dc. Interestingly, the addition of ESSs decreases overall efficiency due to round-trip power losses, which can reduce efficiency by up to 10% in some cases. These findings align with the results in [7], which identify the 48 V DC voltage level as the most efficient solution. Additionally, some research, like [8], suggests a dual voltage level system with 380 V for high voltage and 48 V for low voltage dc loads. However, a universal standard for low voltage dc distribution systems has yet to be established. This is crucial for unifying technical requirements for further development.

The efficiency of the ac-dc rectifier, a crucial component in managing power transfer between the dc building and the utility grid, significantly impacts the overall performance of buildings. For example, rectifying efficiency can drop to as low as 85% during periods of low power demand from the load side. State-of-the-art converters discussed in the literature can achieve efficiencies above 96% [9]. In the market, commercial products typically demonstrate an efficiency range of 92-95% at their nominal peak power [5]. Nevertheless, recent advances in the power electronics industry, particularly with wide bandgap semiconductors, have led to the development of converters with peak efficiencies of 98%. These products are primarily designed for vehicle-to-grid interactions but are also applicable to dc house topologies. Moreover, in recent years, the use of hybrid solutions, featuring both ac and dc distribution networks, has gained research interest [10]. This approach aims to harness the benefits of both systems while minimizing their individual drawbacks if used alone. Fig. 2 demonstrates the collected state-of-the-art efficiency curves of bidirectional isolated ac-dc power converters in the literature [9], [11]-[14].

Despite the higher performance of dc houses in islanded mode compared to ac or hybrid topologies, further research is needed to assess these topologies performance when linked to ac distribution systems. Key research questions for these emerging technologies include: What is the most effective design for internal electricity distribution in residential

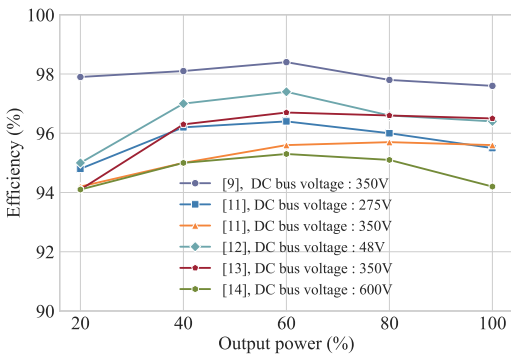


Fig. 2. Efficiency curves of bidirectional isolated ac-dc converters.

buildings? How would a pure dc-based house perform when connected to an ac distribution system? What level of energy efficiency improvement can be achieved by transitioning from an ac to a pure dc house? Are hybrid solutions superior to pure dc networks in the absence of a readily available dc distribution system? And how occupants' consumption pattern could affect the overall performance of the dc house?

In this paper, to address the questions, a detailed analysis of the performance of dc and partial dc houses operating within the current utility grid topology is conducted. For this purpose, the annual consumption profiles of 16 buildings have been studied, assuming their conversion to nZEB using ac, dc, and partial dc technologies. This research, specifically focused on Estonia, aims to assist experts and researchers in understanding the actual performance of these systems in Estonian houses and the broader context of North European climate conditions.

II. METHODOLOGY

A. Transferring Conventional Buildings to nZEBs:

Annual hourly average power consumption data was gathered from a rural district in Estonia. Assuming all buildings require integration of PV and energy storage (ES) systems for transformation into nZEBs, the initial step involved determining the appropriate component sizes. The principles for determining the optimum size configuration are outlined in Algorithm 1. For simplicity, the solar energy generation profile recorded in the same area was uniformly applied to all buildings.

In the size optimization process, economic considerations were excluded. Therefore, the optimal ESS size was determined by plotting the Total Energy Exchange (TEE) curve with the grid, varying the ES size from 1 kWh to 40 kWh. The optimal size is selected at the point where the curve reaches saturation, and further increases in ES size do not yield significant performance improvements.

Algorithm 1 Optimum PV and ES size calculation for nZEBs

```

1: Input: Annual load profiles ( $\bar{l}_p$ ), Annual PV profile ( $p_{pv}$ ).
2:  $P'_{pv} = \text{sum}(p_{pv}) / \text{Max}(p_{pv})$ 
3: for  $i = 1$  to 16 do:
4:    $L_p[i] = \text{sum}(\bar{l}_p[i, :])$ 
4:    $pv_{size}[i] = \text{round up}(\frac{L_p[i]}{P'_{pv}})$ 
5:    $\bar{p}_{pv}^i = pv_{size}[i] * p_{pv}$ 
6:   for  $es$  in  $\vec{ES}$ :
7:     for  $h = 1$  to 8760 do:
8:       if  $\bar{p}_{pv}^i[h] \geq \bar{l}_p[i, h]$ :
9:         Supply load:  $\bar{p}_{pv}^i[h] = \bar{p}_{pv}^i[h] - \bar{l}_p[i, h]$ 
10:        if  $\bar{p}_{pv}^i[h] > 0$ :
11:          Charge es:  $\bar{p}_{pv}^i[h] = \bar{p}_{pv}^i[h] - \bar{es}_{ch}[h]$ 
12:          if  $\bar{p}_{pv}^i[h] > 0$ : Export to utility grid
13:        else:
14:          Supply load:
15:           $\bar{l}_p[i, h] = \bar{l}_p[i, h] - \bar{p}_{pv}^i[h] - \bar{es}_{ch}[h - 1]$ 
16:          if  $\bar{l}_p[i, h] > 0$ : Import from utility grid
17:        end for
18:        Calculate Total exchange energy with grid ( $TEE_i^{es}$ )
19:      end for
20:      for  $i = 1$  to 16 do:
21:        plot  $TEE_i$ 
22:        Select optimum es size for each building
23:      end for
24:    end for
25:  end for
26: Return: optimal setup values  $PV_{size}, ES_{size}$ 

```

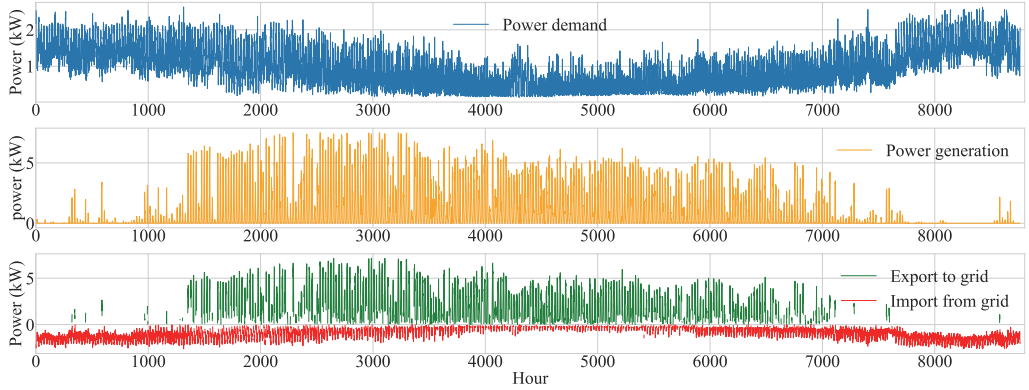


Fig. 3. Annual load profile, renewables power generation, and exchange power with the electricity utility grid.

Additionally, the implemented EMS, obtained from [15], gives priority to PV-generated and ESS-stored energy over the utility grid for meeting energy demands, and exports surplus energy to the grid only when the ESS is full. The Annual power flow and interaction with the grid is depicted in Fig. 3.

B. Power Electronics Specifications:

For estimating dc building performance, the distribution topology shown in Fig. 1.c was considered. The PV MPPT converter and the ES dc-dc bidirectional converter were selected based on the specifications outlined in [16] and [17], respectively. The dc bus voltage was set at 350 V. The ES round trip efficiency (η_R^{es}), excluding the power conversion stages, was selected as 97%. For a hybrid solution a dual-purpose power converter, proposed in [9], is selected to connect PV and ES to the ac and dc buses. In a hybrid solution, as shown in Fig. 1.b it is possible to feed ac and dc loads from both grid and renewable resources.

For each house, the size of the power electronics infrastructure is designed to accommodate the maximum load consumption, maximum PV generation, and a quarter of the ES nominal capacity. This sizing applies to the isolated bidirectional ac-dc, MPPT, unidirectional, and bidirectional dc-dc converters, respectively. In each building, the efficiency curves for the converters have been scaled up or down to match the sizes of selected converters.

C. Energy Loss Analysis

The Energy loss analysis was conducted considering the converters' nominal peak power and efficiency curves and the duration for which they operate at each power level. To this end, the energy loss has been calculated as:

$$E_{loss} = P_{loss} \times t, \quad (1)$$

$$P_{loss} = P_{exchange} \times \eta_{opt}, \quad (2)$$

$$\eta_{opt} = \frac{P_{out}}{P_{in}}, \quad (3)$$

where $P_{exchange}$ is the amount of exchange power with the utility grid, η_{opt} is the converters efficiency in operating power, t is the energy exchange duration, and E_{loss} is the total amount of energy loss during power transformation. P_{out} and P_{in} are converters output and input power, respectively, and P_{loss} is the power loss. Due to the negligible loss in the wires, energy losses in the buildings' distribution systems are disregarded. In dc solutions, only one conversion stage is required to connect the generator,

storage, and dc load to the dc bus. Thus, energy losses are calculated as:

$$E_{loss}^{pv,l} = P_{pv} \times \eta_{opt}^{pv} \times \eta_{load}^{load} \times t, \quad (4)$$

$$E_{loss}^{pv,es} = P_{pv} \times \eta_{opt}^{pv} \times \eta_{opt}^{es} \times t, \quad (5)$$

$$E_{loss}^{es,l} = P_{es} \times \eta_{opt}^{es} \times \eta_{load}^{load} \times \eta_R^{es} \times t, \quad (6)$$

$$E_{loss}^{pv,grid} = P_{pv} \times \eta_{opt}^{pv} \times \eta_{grid}^{grid} \times t, \quad (7)$$

$$E_{loss}^{g,l} = P_{grid} \times \eta_{opt}^{grid} \times \eta_{load}^{load} \times t, \quad (8)$$

where $E_{loss}^{x,y}$ is the energy loss amount when power is transferred from x to y for t time period. η_{opt}^* , $*$ $\in \{pv, es, load, grid\}$ is the efficiency of the PV, ES, grid, and dc load converters in their operating power, respectively. P is the amount of delivered power in all equations. In conventional ac nZEBs topology, the energy losses are calculated as:

$$\tilde{E}_{loss}^{pv,l} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{load}^{load} \times t, \quad (9)$$

$$\tilde{E}_{loss}^{pv,es} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{opt}^{es} \times t, \quad (10)$$

$$\tilde{E}_{loss}^{es,l} = P_{es} \times \tilde{\eta}_{opt}^{es} \times \tilde{\eta}_{load}^{load} \times \eta_R^{es} \times t, \quad (11)$$

$$\tilde{E}_{loss}^{pv,grid} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times t, \quad (12)$$

$$\tilde{E}_{loss}^{g,l} = P_{grid} \times \tilde{\eta}_{opt}^{grid} \times t, \quad (13)$$

where $\tilde{E}_{loss}^{*,*}$ is the energy loss in the conversion stages, between PV, ES, grid, and ac loads. $\tilde{\eta}_{opt}^*$ is the efficiency of dc-ac and ac-dc power converters in their operating point. Finally, if the hybrid solution is considered for nZEB topology, the energy losses are calculated as:

$$\tilde{E}_{loss}^{pv,lac} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{opt}^{lac} \times t, \quad (14)$$

$$\tilde{E}_{loss}^{pv,es} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{opt}^{es} \times t, \quad (15)$$

$$\tilde{E}_{loss}^{pv,grid} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{opt}^{hbd} \times t, \quad (16)$$

$$\tilde{E}_{loss}^{pv,lac} = P_{pv} \times \tilde{\eta}_{opt}^{pv} \times \tilde{\eta}_{opt}^{hbd} \times t, \quad (17)$$

$$\tilde{E}_{loss}^{es,lac} = P_{es} \times \tilde{\eta}_{opt}^{es} \times \tilde{\eta}_{opt}^{lac} \times \eta_R^{es} \times t, \quad (18)$$

$$\tilde{E}_{loss}^{es,lac} = P_{es} \times \tilde{\eta}_{opt}^{es} \times \tilde{\eta}_{opt}^{hbd} \times \eta_R^{es} \times t, \quad (19)$$

$$\tilde{E}_{loss}^{g,lac} = P_{grid} \times \tilde{\eta}_{opt}^{hbd} \times \tilde{\eta}_{opt}^{lac} \times t, \quad (20)$$

where $\tilde{E}_{loss}^{*,lac}$ is the energy loss during the supply of dc loads from grid, PV, or ES. and $\tilde{E}_{loss}^{*,lac}$ is the energy loss during the supply of ac loads from grid, PV, or ES. $\tilde{\eta}_{opt}^{hbd}$ is the efficiency of hybrid dual purpose power converter, which links the buildings' internal ac and dc distribution networks. $\tilde{\eta}_{opt}^{lac}$ is the efficiency of dc loads' dc-dc converter. Furthermore, the efficiencies of the internal power conversion for ac and dc loads are extracted from [18], [19], [20] and is collected in TABLE I.

III. RESULTS AND DISCUSSIONS

A. Power Electronics Size Optimization Outcomes

Selected infrastructure sizes have been collected in TABLE II. The TEE curves for some houses if operating in ac and dc mode, with various ES sizes, are presented in Fig. 4. As observed, increasing the ES size reduces the power exchange ratio. Furthermore, the energy exchange ratios for dc topologies are less than ac systems. Later we will explore this figure in more detail.

TABLE I
HOME APPLIANCES' POWER CONVERSION EFFICIENCIES

Appliance	Efficiency (%)		Appliance	Efficiency (%)	
	dc-dc	ac-dc		dc-dc	ac-dc
Laptop	93.2	85.9	Lights	99	96
Refrigerator	95	86	TV	98	90
Mixer grinder	98	96	Oven	97.2	95.6
Vacuum cleaner	95	93	Heating	99.9	99.8

TABLE II
POWER ELECTRONICS SIZE SELECTION FOR HOUSES

House code	Grid tied ac-dc converter	PV size	ES size	dc loads converter
1	3 kW	5 kW	20 kWh	2 kW
2	2.5 kW	2.5 kW	11 kWh	1.5 kW
3	4 kW	3 kW	15 kWh	3.5 kW
4	1.5 kW	2 kW	4 kWh	1 kW
5	3.5 kW	4.5 kW	17 kWh	2.5 kW
6	4 kW	1.5 kW	8 kWh	3.5 kW
7	4.5 kW	6 kW	13 kWh	3 kW
8	5 kW	6.5 kW	16 kWh	3 kW
9	4 kW	4.5 kW	12 kWh	2.5 kW
10	1 kW	1 kW	5 kWh	1 kW
11	6.5 kW	4.5 kW	13 kWh	4 kW
12	6 kW	5 kW	12 kWh	4 kW
13	3.5 kW	3 kW	10 kWh	2.5 kW
14	3.5 kW	4.5 kW	12 kWh	2.5 kW
15	2.5 kW	2.5 kW	8 kWh	1.5 kW
16	3 kW	7.5 kW	20 kWh	2 kW

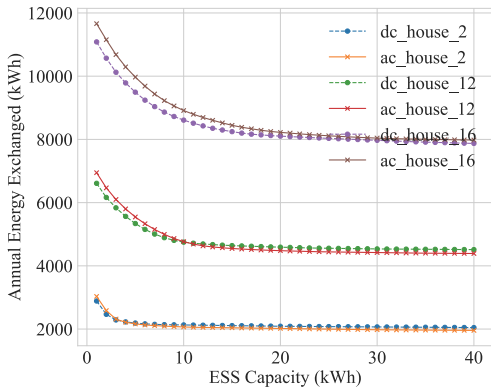


Fig. 4. ES size impacts on annual energy exchange for some ac and dc buildings.

B. Device Operating Region Analysis

Considering that converters do not consistently operate at peak efficiency, the time spent in low power modes significantly affects the system's overall performance. Fig. 5 displays the duration of the main isolated bidirectional ac-dc converter operating at different capacity levels over the year for each dc building. It can be observed that in all the studied cases, with appropriately sized PV and ES systems, the main converter predominantly operates in the low

power, which corresponds to less than a quarter of its nominal peak power. This power capacity is highlighted in red to emphasize the higher loss values. These losses are a result of the lower efficiency of converters at this reduced power level. In most of the studied cases, the main grid-tied converter operates for less than 5% of the total operation period at above 75% of its nominal peak power. Furthermore, on average, the converters transfer power at less than half their capacity for 90% of the operation periods. These results question the overall efficiency gains achievable through converting an ac building into a dc building. However, it is important to note that these results are case-sensitive and influenced by factors such as the user's lifestyle, the energy grade of the buildings, and the area's climate conditions. Despite these variables, the data can still provide a realistic insight into the real potential of dc transformation in Estonian households. Furthermore, to draw a more comprehensive analysis and conclusion, a comparison of the performance and energy losses of ac, dc, and hybrid nZEBs is also necessary.



Fig. 5. Share of each operation power for the main isolated bidirectional ac-dc converter in the buildings. (colors show the operating power level).

C. Distribution Topologies Performance Comparison

To compare the performance of different topologies, the total energy loss for each topology across all studied cases has been analyzed and compared. Fig. 6. presents a comparison of the total annual renewable energy loss for all houses, highlighting the differences when the solar power converter is connected to ac versus when it is connected to dc distribution networks. So, on average, 52% of renewable energy loss could be reserved in both dc and hybrid solutions, in comparison to conventional ac topology.

Fig. 4 reveals that the annual energy exchange with the utility grid for a dc house is less than conventional ac houses. However, energy exchange in conventional ac houses is almost lossless compared to dc houses. This is due to the required dc-ac conversion stage needed to connect a dc building to an ac distribution network. Therefore, for a comprehensive performance analysis, total power conversion losses in different topologies must be evaluated. Since hybrid or partial dc topologies incorporate both ac and

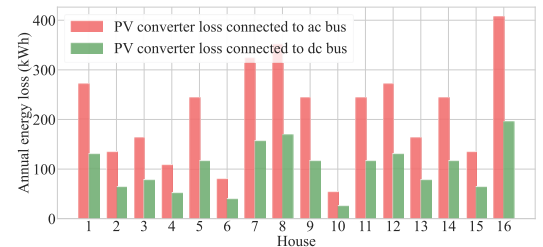


Fig. 6. PV power converters' annual energy losses when connected to ac bus and dc bus.

dc buses within the buildings, determining the capacity of each bus is critical to minimize power flow between the internal ac and dc buses. In this study, 40% of the load demand was considered as ac, with the remainder supplied through a dc bus. This ratio was chosen through trial and error, and further research into hybrid topologies is necessary to identify the optimum balance.

Fig. 7 compares the annual total energy losses due to power conversions in all topologies and houses. It shows that the nZEB using ac technology incurs the highest losses. Conversely, the hybrid solution demonstrates the best performance in terms of energy loss. This outcome is expected and reasonable, as the hybrid solution has the advantage of directly feeding high-demand ac loads from the grid, bypassing the power converter. Additionally, as shown in Fig. 5, in dc buildings, the interlink isolated ac-dc converter mostly operates in the first quarter of its maximum capacity. Consequently, this leads to lower efficiency and higher conversion losses.

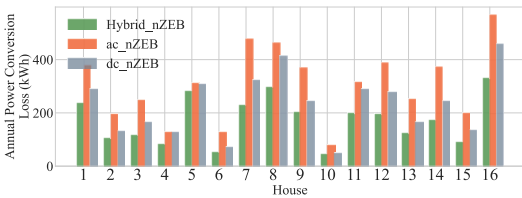


Fig. 7. Power conversion losses comparison in hybrid, ac, and dc topologies.

Fig. 8 compares grid interactions for all topologies. The hybrid solution shows lower performance compared to both dc and ac topologies, with a higher rate of annual energy trade with the main utility grid. The effectiveness of the hybrid solution heavily relies on the implementation of a suitable EMS. In our study, the supervisory algorithm lacks forecasting capabilities, thus failing to optimally manage stored energy for both dc and ac loads. Consequently, the ES system is dominantly used for powering dc loads, leading to increased energy exchanges and imports from the grid for feeding ac loads. Therefore, without an advanced EMS capable of efficiently controlling ES and power distribution, the hybrid solution might not consistently achieve optimal performance.

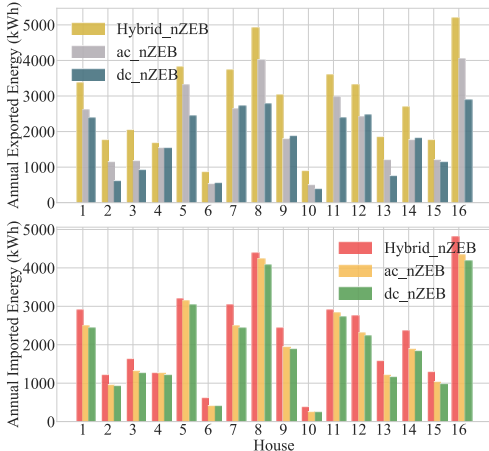


Fig. 8. Annual energy trade comparison for hybrid, ac, and dc topologies in all studied buildings.

D. Buildings Consumption Pattern Comparison

For dc technology, a promising approach is the formation of energy districts by integrating nearby buildings or houses. This facilitates power transfer among prosumers and enables energy exchanges independent of the main ac utility grid. Such a configuration could enhance the overall performance of dc buildings and minimize power interactions with the ac grid. However, for optimal functioning, it is crucial to analyze electricity generation and consumption patterns in each building to identify potential overlaps. For instance, solar energy generation occurs simultaneously for all nearby users. This leads to concurrent availability of surplus energy. Additionally, houses' energy consumption patterns often align due to similar lifestyles.

We analyzed daily consumption patterns in all studied cases to assess the feasibility of the establishment of dc energy neighborhoods. For each building, demand was categorized into three sections: 'Low', 'Normal', and 'High'. This classification involved calculating the mean and standard deviation for each case. Consumption values lower than mean minus half the standard deviation were labeled as 'Low', and those higher than mean plus half the standard deviation as 'High'. Values falling outside these ranges were considered 'Normal'. The results of this data analysis are detailed and displayed in Fig. 9.

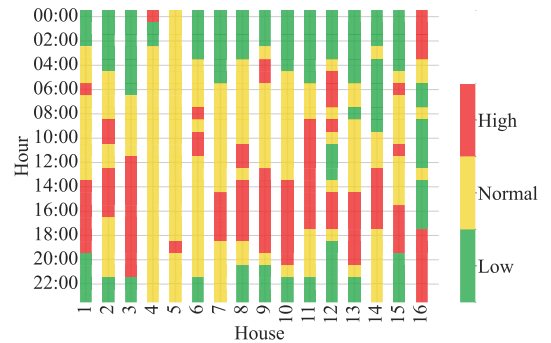


Fig. 9. Power consumption pattern for all studied houses during a day.

Analyzing the consumption habits reveals that most of the buildings, except for houses number 4 and 5, show remarkably similar usage patterns, primarily consuming power during early morning and evening times. However, the pattern for house number 16 is an outlier in the examined buildings. This deviation is due to the presence of an electric vehicle (EV) in this household, which is mostly charged during off-peak night hours. Consequently, house number 16 has high power demand at times when other buildings are experiencing their low-demand phases.

Given that all buildings are equipped with a similar EMS and have adopted solar power as a renewable energy source, it can be concluded that while the concept of constructing local dc neighborhoods for energy trading is appealing, crucial challenges such as overlapping consumption phases and concurrent access to renewable sources will pose difficulties for the system's operation, and reliability. For instance, in the studied neighborhood, the only house that could potentially benefit from energy sharing is house number 16. This is due to it owning an EV and requiring power during midnight hours. So, if other houses also adopt EVs, similar challenges in the local dc distribution network are likely to emerge soon.

E. Discussion

Investigations into various nZEB topologies have revealed significant challenges in implementing a pure dc topology. The anticipated performance improvements are not as substantial as initially expected. This can be attributed to two main reasons. First, our data profiles shows that the interlink ac-dc power converter often operates at its least efficient stage due to low power demands during off-peak periods. Second, during Estonia's winters, with limited solar power availability, the system heavily relies on power from the ac utility grid, leading to energy loss during power conversion. In contrast, hybrid or partial solutions can mitigate this issue by utilizing the ac distribution network to directly supply heavy ac loads, potentially resulting in lower energy losses and higher annual efficiency.

However, the main challenge for hybrid solutions lies in the complexity of the supervisory management algorithm. This algorithm needs to efficiently distribute power between dc and ac loads to outperform conventional ac nZEBs. This complexity demands advanced management algorithms and greater computational resources. Given the current absence of a low voltage dc distribution system, it is evident that both dc and hybrid topologies need further enhancements in power electronics and information technology to compete effectively with conventional ac nZEBs topology.

IV. CONCLUSION

In this study, we investigated the transformation of conventional ac houses into nZEBs, considering three different topologies: ac, hybrid, and dc. The comparison of these topologies shows that in terms of power loss metrics, the hybrid or partial dc solution performs better. In contrast, the dc topology demonstrates the lowest grid interaction, hence the highest energy autonomy compared to the others. While both the hybrid and dc topologies outperform the conventional ac nZEB topology, the improvements are not as significant as anticipated. These findings underscore the need for further development of these alternative technologies to gain market acceptance. Additionally, analyzing consumer consumption patterns reveals that forming dc energy communities with nearby dc buildings may not function as expected due to similar consumption patterns. However, this study's outcomes depend on the specific data and location analyzed. While the results may vary for different users and locations, they offer valuable insights into potential challenges and areas for improvement in dc and hybrid nZEB technologies.

ACKNOWLEDGMENT

This work was supported by the Estonian research council grants PRG675, EAG234 and PRG1463, and the European Union's Horizon Europe research and innovation program under the grant agreement No. 101120657, project ENFIELD (European Lighthouse to Manifest Trustworthy and Green AI).

REFERENCES

- [1] K. Garbesi, "Catalog of dc appliances and power systems," 2012.
- [2] C. Moscatiello, R. Loggia, G. di Lorenzo, A. L. Palma, M. Kermani, R. Faranda, F. Oliva, E. Tironi, R. Araneo, and L. Martirano, "A new proposal for power sharing in lvdc energy community microgrids," *IEEE Transactions on Industry Applications*, 2023.
- [3] B. Glasgo, I. L. Azevedo, and C. Hendrickson, "How much electricity can we save by using direct current circuits in homes? understanding the potential for electricity savings and assessing feasibility of a transition towards dc powered buildings," *Applied Energy*, vol. 180, pp. 66–75, 2016.
- [4] E. L. Carvalho, A. Blinov, A. Chub, P. Emiliani, G. de Carne, and D. Vinnikov, "Grid integration of dc buildings: Standards, requirements and power converter topologies," *IEEE Open Journal of Power Electronics*, vol. 3, pp. 798–823, 2022.
- [5] A. H. Sabry, A. H. Shallal, H. S. Hameed, and P. J. Ker, "Compatibility of household appliances with dc microgrid for pv systems," *Heliyon*, vol. 6, no. 12, 2020.
- [6] S. Habibi, R. Rahimi, P. Shamsi, and M. Ferdowsi, "Efficiency assessment of a residential dc nanogrid with low and high distribution voltages using realistic data," in *2021 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2021, pp. 574–579.
- [7] K. Siraj and H. A. Khan, "Dc distribution for residential power networks a framework to analyze the impact of voltage levels on energy efficiency," *Energy Reports*, vol. 6, pp. 944–951, 2020.
- [8] D. L. Gerber, V. Vossos, W. Feng, C. Marnay, B. Nordman, and R. Brown, "A simulation-based efficiency comparison of ac and dc power distribution networks in commercial buildings," *Applied Energy*, vol. 210, pp. 1167–1187, 2018.
- [9] T. H. Shahsavari, N. V. Kurdkandi, O. Husev, E. Babaei, M. Sabahi, A. Khoshkbar-Sadigh, and D. Vinnikov, "A new flying capacitor-based buck-boost converter for dual-purpose applications," *IEEE Journal of Emerging and Selected Topics in Industrial Electronics*, vol. 4, no. 2, pp. 447–459, 2023.
- [10] Zhang, S. Buso, and T. Caldognetto, "A flexible energy gateway for hybrid nanogrids," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 10, no. 5, pp. 5717–5726, 2022.
- [11] Y. Zhang, W. Song, W. Hua, G. Li, Y. Qian, Z. Yu, H.-Y. Tsai, Y.-M. Chang, and J. Hu, "A boost-inductorless electrolytic-capacitorless single-stage bidirectional isolated ac-dc converter," *IEEE Transactions on Power Electronics*, vol. 38, no. 4, pp. 5469–5478, 2022.
- [12] T. Chen, R. Yu, and A. Q. Huang, "A bidirectional isolated dual-phase-shift variable-frequency series resonant dual-active-bridge gan ac-dc converter," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 4, pp. 3315–3325, 2022.
- [13] E. L. Carvalho, A. Blinov, P. Emiliani, A. Chub, and D. Vinnikov, "Three-phase bidirectional isolated ac-dc matrix-converter with full soft-switching range," *IEEE Access*, vol. 11, pp. 119 270–119 283, 2023.
- [14] S. K. Dube, R. Nair, and P. Das, "Analysis and design of an integrated bidirectional three-phase ac-dc resonant converter," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 5, pp. 4369–4379, 2022.
- [15] H. N. Hokmabad, O. Husev, J. Belikov, D. Vinnikov, and E. Petlenkov, "Energy storage and forecasting error impact analysis in photovoltaic equipped residential nano-grids," in *2023 IEEE 17th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, pp. 1–6, 2023.
- [16] O. Husev, D. Vinnikov, C. Roncero-Clemente, A. Chub, and E. Romero Cadaval, "Single-phase string solar qzs-based inverter: Example of multi-objective optimization design," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 3120–3130, 2020.
- [17] V. Sidorov, A. Chub, D. Vinnikov, and A. Lindvest, "Novel universal power electronic interface for integration of pv modules and battery energy storages in residential dc microgrids," *IEEE Access*, vol. 11, pp. 30 845–30 858, 2023.
- [18] A. Santos, G. Duggan, S. Frank, D. Gerber, and D. Zimmerle, "Endpoint use efficiency comparison for ac and dc power distribution in commercial buildings," *Energies*, vol. 14, no. 18, p. 5863, 2021.
- [19] G. Makarabbi, V. Gavade, R. Panguloori, and P. Mishra, "Compatibility and performance study of home appliances in a dc home distribution system," in *2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*. IEEE, 2014, pp. 1–6.
- [20] O. Luc'ia, I. Cvetkovic, H. Samago, D. Boroyevich, P. Mattavelli, and F. C. Lee, "Design of home appliances for a dc-based nanogrid system: An induction range study case," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 4, pp. 315–326, 2013.

Publication VII

H. N. Hokmabad, O. Husev, D. Vinnikov, J. Belikov, and E. Petlenkov, 'Day-ahead PV Output Power Forecasting Utilizing Boosting Recursive LightGBM-LSTM Framework', in *2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2023, pp. 1–5.

Day-ahead PV Output Power Forecasting Utilizing Boosting Recursive LightGBM-LSTM Framework

Hossein Nourollahi Hokmabad
Oleksandr Husev, Dmitri Vinnikov
Department of Electrical Power
Engineering and Mechatronics,
Tallinn University of Technology,
Tallinn, Estonia

Juri Belikov
Department of Software Science,
Tallinn University of Technology,
Tallinn, Estonia

Eduard Petlenkov
Department of Computer Systems,
Tallinn University of Technology,
Tallinn, Estonia

Abstract—Recently, global photovoltaic (PV) system installations have surged. Precise forecasting is vital for their grid integration and carbon emission cuts. However, due to fluctuating solar radiation, predicting PV output is difficult. Machine learning models, notably Long Short-Term Memory (LSTM) networks, offer a solution. This study presents a novel framework using a boosted recursive Light Gradient Boosting Machine (LightGBM)-LSTM network to forecast daily PV output at hourly intervals. Conducted in Northern Europe—a region with significant solar radiation variability—the models trained on meteorological and historical data showed a 12% improvement in RMSE, a 13% reduction in MAPE, and a 5% increase in the R^2 score compared to standalone LSTM models.

Index Terms—Solar power forecasting, numerical weather prediction, LightGBM, LSTM, boosting ensemble method.

I. INTRODUCTION

Renewable resources are essential for achieving carbon neutrality and sustainable energy. However, due to their intermittent nature, they aren't as consistent as dispatchable power plants. Without precise short-term predictions of renewable availability, utility companies may overproduce energy, potentially challenging grid stability. Furthermore, PV setup owners entering the energy market without adequate estimation of their achievable production level may incur financial penalties for under or over-delivering. Moreover, in the small-scale size installation, for instance, commercial or residential real estates, accurate projection of available renewable energy boosts the energy management systems performance as well [1]. Therefore, precise forecasting of both short and long-term energy generation is critical for achieving environmentally sustainable grid management and ensuring profitability in the energy market [2].

Lately, within the forecasting realm, Machine Learning (ML) and Deep Learning (DL) techniques are becoming predominant in state-of-the-art applications [3]. Among ML/DL techniques, hybrid methods that combine results from different methods have shown superiority over simpler standalone methods. Attention based convolutional neural networks (CNN) model has been adopted in [4] for ultra short-term predictions. LSTM models, as an improved version of recurrent neural networks, have been merged into many hybrid models in the literature. For instance, a composite of CNN-LSTM and multi-layer perceptron networks integrated with error correction and decomposition methods has been proposed in [6] for step-by-step solar irradiance forecasting. In [7], Coati optimization algorithm has been applied with LSTM-CNN framework to perform wind and solar power forecasting.

Regression models, adept at uncovering hidden relationships between target variables and features, have reliably served the solar forecasting sector. Notable

implementations include the Support Vector Regressor [8], Gradient Boosted Regression Tree [9], and Weighted Gaussian Process Regression [10] for short-term forecasting. Some researchers have amplified the efficacy of these models by integrating them with classification [3] or transfer learning methods [11]. Given their inherent structure, tree-based regression models are particularly favored for non-linear regression tasks. The LightGBM model, an ensemble tree-based regression approach, has been applied in forecasting loads [12], PM 2.5 ranges [13], and PV power [14].

Building on these insights and established methodologies in the field, our research aims to push the boundaries of day ahead PV forecasting. In this study, a novel approach for day-ahead output power forecasting of a PV setup is proposed, combining the LightGBM and LSTM cells. Unlike previous studies, this approach aims to harness the strengths of both tree-based ensemble regression and time series regression analysis. A boosting approach merges the LightGBM model, capturing complex weather-to-power relationships, with LSTM cells that learn temporal data dependencies for accurate recursive multistep forecasts. The integration of a meta-learner and tuning stages further refines predictions, enhancing the combined model's performance and offering more precise day-ahead forecasts.

II. PROPOSED METHODE AND DATA DESCRIPTION

As previously mentioned, this study proposes an ensemble approach. Ensemble models have demonstrated superior performance in forecasting tasks compared to conventional statistical and base ML models [15]. Fig. 1 provides an abstract structure of the suggested framework.

In this pipeline, the LightGBM model is trained with historical time series output power records from the target PV site, along with meteorological measurements of the located area. Simultaneously, LSTM units have been trained on the same dataset to learn temporal features of the target data. During the forecasting stage, numerical weather prediction (NWP) data for the target period (next 24 hours) is utilized to

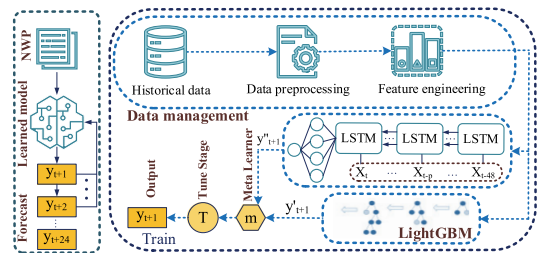


Fig. 1. An abstract structure of the proposed LightGBM-LSTM framework. Here, X is a historical time series data, y' is the validation stage output, and y is a predicted target value, m is a meta learner, and T is a tune stage.

This work was supported by the Estonian Research Council grants PRG675 and PRG1463.

obtain initial results from the regressor (y''_{t+1}). The corresponding regression output along with an output of the stacked LSTM network (y'_{t+1}) have been utilized to train a gradient boosting regression (GBR) meta-learner to enhance the final prediction value (y_{t+1}). Finally, recursive forecasting method utilizing sliding window technique is applied to leverage the past forecasting values to generate the next set of predictions. However, recursive multi step timeseries forecasting techniques are prone to error propagation and show a poor performance in the long-range forecasting, but the proposed auxiliary regression, meta learner, and tuning tools try to enhance the performance of the main framework.

A. LightGBM

LightGBM [16] is a ML framework developed by Microsoft, which is open-source and implements the Gradient Boosting Decision Tree (GBDT) algorithm. Despite its ability to handle big data, LightGBM does not sacrifice accuracy, making it a powerful and practical ML tool. To improve its efficiency when dealing with big data, such as datasets with thousands of features and millions of instances, LightGBM incorporates two techniques: Gradient-Based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) [18]. The EFB technique groups exclusive features, such as one-hot-encoded features, to reduce dimensionality and improve model training speed. The GOSS technique involves discarding a significant portion of data instances with small gradients and using only those with larger gradients to estimate information gain [18]. This approach leads to a more streamlined and efficient computation of information gain, which is a metric used in decision tree algorithms to measure the relevance of features in predicting the target variable and discovering optimal split points. Information gain is also widely used in feature selection processes.

B. LSTM

LSTM networks [17] are a variant of recurrent neural networks (RNNs) that can model long-term dependencies by employing layers of LSTM cells. Unlike traditional RNNs, LSTMs are robust against vanishing and exploding gradient problems, making them well-suited for retaining information over long-term sequences. Within the LSTM unit, a cell state is modified by two gates, the forget and input gates, and an output gate generates a hidden state signal based on the current cell state. Multi stacked LSTM units enables the network to learn temporal and sequential dependencies of input data. In this study (as shown in Fig. 2), the LSTM network was trained on 48 hours of past data for one-step prediction, followed by a recursive strategy for subsequent predictions.

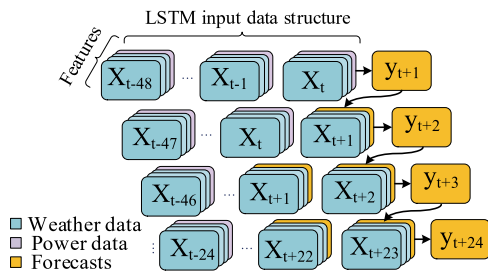


Fig. 2. LSTM network data structure.

C. Meta Learner

Meta-learning techniques can be employed to refine prediction accuracy further. In this context, we've chosen the

GBR [18] method as our meta-learner. GBR stands out as a formidable tree-based ensemble regression technique, which operates by sequentially building base learners to minimize the residuals of preceding stages. These base learners, by rectifying each other's errors, collectively offer more dependable outcomes in regression analysis.

D. Tune Stage

The final tuning units in this study perform feasibility checks on predicted outcomes based on the technical characteristics of the PV farm, which range from 0-25 kW. Predictions that exceed the maximum or minimum output power of the system are adjusted to the maximum or zero, respectively. Additionally, non-zero outcomes corresponding to twilight times are set to zero using sunrise and sunset times from the corresponding calendar day of the previous year. These corrections during the ML model development process can improve both the performance and reliability of the obtained results.

E. Site Location and Data Description

Tallinn, located in the northeastern region of Europe, experiences imbalanced levels of solar irradiance throughout the year. Fig. 3 presents a comparative analysis of the total daily surface net solar radiation between Tallinn and several European cities. The analysis reveals that Tallinn experiences a greater degree of variability in solar radiation levels compared to most of these cities excluding Oslo, rendering the task of accurate model training more challenging.

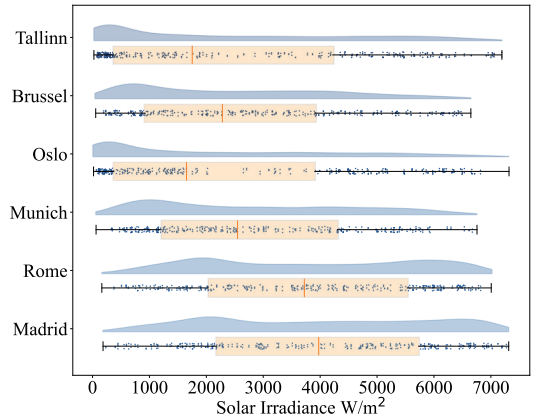


Fig. 3. Comparison of daily accumulated solar irradiance at multiple European cities during 2010 to 2022.

PV output power data was collected from a small-scale PV installation located within 20km of Tallinn, Estonia over the years 2021 and 2022. The present data consists of compiled information regarding farm generated power, recorded and accumulated at 1-minute intervals. To align the granularity of the output power data with the meteorological dataset, the data was down-sampled from 1-minute intervals to hourly intervals and aggregated accordingly.

To obtain archived meteorological data, this study utilizes the ERA5 reanalysis dataset, which is publicly available and provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) [19]. This dataset offers high spatio-temporal resolution, with a resolution of 31km and 1 hour, respectively. It includes detailed records of various climate parameters such as solar irradiance, temperature, dew point,

wind direction, wind speed, cloud cover, cloud layers, visibility, precipitation amount, snow cover, and others. These records span from 1959 onwards, providing a rich source of meteorological data for the study. Moreover, in the forecasting stage, weather parameters data for the subsequent day was procured by utilizing freely accessible NWP application programming interfaces (APIs).

F. Feature Engineering

Various feature selection methods, such as Pearson correlation coefficient (PCC), Recursive Feature Elimination (RFE) with Random Forest Regressor, and feature importance ranking with Ridge regression were used to eliminate redundant and highly correlated features that may negatively impact model performance. These methods are suitable for continuous data and regression analysis. The feature selection process reduced the number of features from 44 to 11, indicating the effectiveness of the selected methods in identifying the most relevant meteorological features for PV output power forecasting. However, it should be noted that the specific number of features selected may vary depending on the nature of the data and the particular feature selection methods used.

To improve the performance of the proposed model, additional features were generated and added to the existing dataset. These features included a daytime index, which was produced using the clear sky surface net solar radiation value. Given that the PV output power consistently registers as zero in the absence of sunlight, consequently, during the training of the regression model, solely those rows exhibiting a daytime value of one have been utilized. In addition, month and hour values were extracted from the timeseries index for each instance, enabling the model to capture seasonal and diurnal patterns in the dataset.

Furthermore, the PCC values among the target column and its lags were examined, revealing strong correlations between current PV output and its time shifts. To incorporate this temporal information into the model, time lags with PCC values higher than 0.75 (including 1, 2, 24, 25, 48, and 72 hours) were added to the dataset. It is worth mentioning that snow-related features, such as coverage, albedo, and layer thickness, were included in the dataset to account for their impact on PV output power generation during autumn and winter. These features are location-specific and play a crucial role in accurate forecasting during the colder months. In particular, snow coverage and layer thickness can affect the amount of sunlight absorbed by the solar panels, while albedo (the reflective properties of the snow surface) can impact the amount of solar radiation reflected or absorbed by the panels. By including these location-specific features, the proposed model can better capture the seasonal variability.

G. Evaluation Metrics

In the realm of evaluating ML models' performances, several metrics have been identified. Here, the Root Mean Square Error ($RMSE = \sqrt{\frac{1}{n} \sum_{h=1}^n (y_h - y'_h)^2}$), Mean Absolute Percentage Error ($MAPE = \frac{\%100}{n} \sum_{h=1}^n \left| \frac{y_h - y'_h}{y_h} \right|$), and coefficient of determination ($R^2 = 1 - \frac{SS_{res}}{SS_{total}}$), have been chosen as measures to gauge the accuracy and reliability of the models. Where n is number of forecasting steps (here considered as twenty-four) and h is a current forecast step, y

and y' are actual and forecasted values, respectively. Also, $SS_{res} = \sum_{h=1}^n (y_h - y'_h)^2$ and $SS_{total} = \sum_{h=1}^n (y_h - \bar{y})^2$ are sum of squared residuals and the total sum of squares, respectively.

H. Baselines

The proposed method has been evaluated against single established statistical models within the time series forecasting and regression analysis fields. The selected models for comparison include Persistence, LSTM, LightGBM, and XGBoost [20]. Each of the benchmark models has been optimized to exhibit their optimal performance to ensure a fair comparison. The Persistence model utilizes a straightforward methodology for forecasting, whereby it assumes that the predicted value for PV power is equivalent to the value from the corresponding hour of the previous day.

III. MODEL STRUCTURE AND RESULTS

The model advancements were executed using the Keras 2.10 library in Python 3.10.7 and implemented on a desktop computer featuring an Intel Core-i7-7700K CPU (4.20GHz) and a 32 GB memory capacity. Furthermore, hyperparameters optimization process have been conducted utilizing Optuna [21] automatic hyper parameter optimization framework and Google Colab cloud service.

A. Model Hyperparameters

During LightGBM model design process, various hyperparameters should be calibrated to assure the model performance. These hyperparameters can be classified into three categories: data-related, boosting-related, and regularization-related. Data-related hyperparameters include parameters such as the maximum depth of a tree, minimum number of samples in each leaf node, and the fraction of features to consider when building a tree. Boosting-related hyperparameters include the number of trees to be used in the model, learning rate (LR), and subsample rate. Regularization-related hyperparameters include parameters such as L1 and L2 regularization. The selected hyperparameters for regression model have been collected in Table I.

TABLE I
LIGHTGBM HYPER PARAMETERS

Hyper parameter name	Value/Name	Range
Boosting type	Goss	GOSS-DART
Learning rate	0.0098	1e-4-3e-3
Feature fraction	0.7	0-1
Number of leaves	61	20-100
Minimum child samples	24	1-40
Minimum child weight	0.0024	1e-4-1e-1
Regression alpha (L1)	0.0029	1e-4-1e3
Regression lambda (L2)	0.0013	1e-4-1e3
Number of estimators	972	200-1500
Subsample	0.568	0-1

Furthermore, stacked LSTM network has been optimized utilizing grid search technique. This technique guarantees finding the best answer in the search space, however, is computationally expensive and slower than other methods. Specifically, the hyperparameters optimized for the stacked LSTM network included the number of LSTM units, the learning rate, activation function type, the dropout rate, and number of neurons in the hidden dense layer. Table II tabulates these values. Additionally, the training data for

LSTM model was obtained from the last two days' samples, corresponding to 48 data rows, to predict one step ahead value.

TABLE II
LSTM NETWORK CHARACTERISTICS

Hyper parameter name	Value/Name	Range
LSTM layer 1 units	41	20-100
Activation function	Tanh	Relu, Tanh
Drop out	0.1	0.05-0.25
LSTM layer 2 units	55	20-100
Activation function	Tanh	Relu, Tanh
Drop out	0.25	0.05-0.25
Hidden dense layer neurons	131	50-200
Activation function	Relu	Relu, Tanh
Output dense layer neurons	1	1
Optimizer	RMSprop	RMSprop/Adam
rho	0.9	-
Learning rate	0.0024	1e-3-3e-2
Loss metrics	MSE	-

B. Results and Discussions

In this section first the result of each sub-model will be reviewed then they will be compared to the main ensemble model performance. Also, the outcomes are benchmarked with baselines to assess the extent to which combining the two models enhances forecasting accuracy. It is important to note that the quality and quantity of data can have a significant impact on the performance of each model, and thus, each model's performance should be evaluated based on the dataset it was trained on. Fig. 4 displays the prediction versus actual values for four ML models: LightGBM, 1-step LSTM, multistep recursive LSTM, and XGBoost. Among these models, the LightGBM model demonstrated the best performance, while XGBoost had a similar R^2 score but was more computationally expensive. Additionally, the 1-step LSTM model outperformed the multistep recursive LSTM model due to its step-by-step update during the forecasting process. Moreover, LightGBM was found to be an effective model for boosting the multistep recursive LSTM's performance, as it delivered excellent results while reducing computational costs compared to XGBoost model.

Data leakage is a prevalent issue in time series forecasting that can result in a model's overfitting and inflated performance metrics. This occurs when future instances, which should be unseen during LSTM model training, are included in the regression models training set. As a result, the model's validation accuracy is artificially higher than its performance on unseen data. To avoid data leakage during the boosting stage of the LSTM model, the LightGBM model was trained using an unshuffled train/test split data. While this approach reduced the Light-GBM model's R^2 score from 0.93 to 0.90, it ensured that the LSTM model's performance was not impacted by data leakage issues. This precautionary measure underscores the importance of carefully handling data in time series forecasting to develop robust and accurate models. The results of the proposed model and the benchmark models (LightGBM, 1 step LSTM, and multi-step LSTM) are presented in Fig. 5, depicting the forecasting accuracy over a one-week duration.

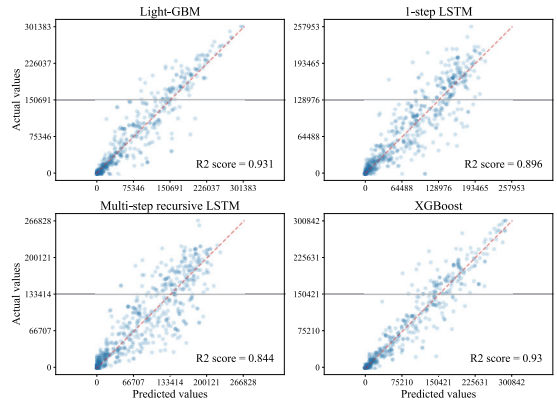


Fig. 4. Comparison between predicted values and actual values in four distinct ML models.

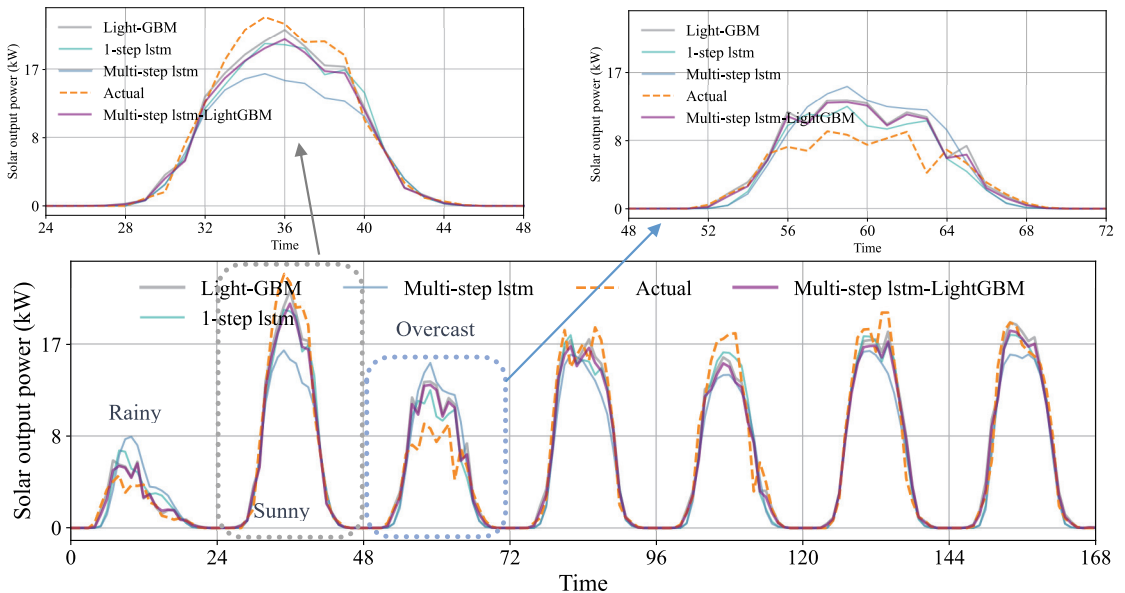


Fig. 5. Forecasting outcomes of proposed model and benchmark models during a week with all types of weather conditions.

The figure illustrates that all models are capable of accurately capturing the sophisticated relationships between the input features and the PV output power. However, during overcast and partly cloudy days, the models' forecasting accuracy is challenged by the randomness of cloud movements. On the other hand, for fully or mostly sunny days, all algorithms have shown satisfying forecasting results. Table III summarizes the performance metrics of the proposed model and benchmark models. Boosting the multi-step LSTM model with LightGBM improved the R^2 score from 0.844 to 0.932 and decreased the MAPE and RMSE values by 27% and 25%, respectively. Additionally, the multi-step LSTM-LightGBM model outperformed the 1-step LSTM model, indicating the strength of regression models when the forecasting weather conditions and parameters are accurate enough for the corresponding day. However, it is worth mentioning that the LightGBM model has demonstrated promising performance, even without utilizing the merging stage with the LSTM network. This suggests that LightGBM can be a reliable standalone model for PV output power forecasting providing enough data in the learn stage.

TABLE III
MODEL PERFORMANCE BENCHMARKS

Model	R^2 score	MAPE%	RMSE (Wh)
Proposed	0.932	17.82	1586.79
1-step LSTM	0.896	20.61	1791.65
Multi-step recursive LSTM	0.844	24.62	2142.03
LightGBM	0.904	19.58	1693.05
XGBoost	0.898	20.49	1764.18
Multi-step persistence	0.593	34.73	3648.72
1-step Persistence	0.742	29.31	2869.86

IV. CONCLUSION

In conclusion, this study presents a novel approach for day-ahead output power forecasting of a PV farm by integrating the LightGBM and LSTM models using a meta learner and tuning units. The proposed method surpasses individual LightGBM, LSTM, and other benchmark models, exhibiting an 11% improvement in the R^2 score and 13% reduction in MAPE compared to conventional multi-step LSTM networks. Although the tuning strategy, feature engineering, and other approaches have demonstrated improved final model performance, the proposed framework has weaknesses, particularly during challenging weather conditions such as partly cloudy and overcast days. Furthermore, it is computationally expensive due to utilizing LSTM networks alongside the meta learner. Future work could include evaluating the model's performance in different locations and exploring more advanced methods and strategies for the tune unit.

ACKNOWLEDGMENT

This work was supported by the Estonian Research Council grants PRG675 and PRG1463.

REFERENCES

- [1] Hokmabad, H. N., Belikov, J., Husev, O., Vinnikov, D., "State-of-the-art activity recognition and prediction techniques applicable to the home energy management system," *IEEE 7th International Energy Conference*, pp. 1-7, 2022.
- [2] Tawn, R., and J. Browell. "A review of very short-term wind and solar power forecasting," *Renewable and Sustainable Energy Reviews*, vol.153, pp. 111758, 2022.
- [3] M. Nejadi and N. Amjadi, "A new solar power prediction method based on feature clustering and hybrid-classification-regression forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1188-1198, 2021.
- [4] Jalali, S.M.J., Ahmadian, S., Nakisa, B., Khodayar, M., Khosravi, A., Nahavandi, S., Islam, S.M.S., Shafie-khah, M. and Catalão, J.P., "Solar irradiance forecasting using a novel hybrid deep ensemble reinforcement learning algorithm," *Sustainable Energy, Grids and Networks*, vol.32, p.100903, 2022.
- [5] A. Shadab, S. Ahmad, and S. Said, "Spatial forecasting of solar radiation using ARIMA model," *Remote Sensing Applications-Society and Environment*, vol. 20, 2020.
- [6] Liu, J., Huang, X., Li, Q., Chen, Z., Liu, G. and Tai, Y., "Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD," *Energy Conversion and Management*, vol. 280, p.116804, 2023.
- [7] Abou Houran, M., Bukhari, S.M.S., Zafar, M.H., Mansoor, M. and Chen, W., "COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications," *Applied Energy*, 349, p.121638, 2023.
- [8] U. K. Das, K. S. Tey, M. Y. I. Bin Idris, S. Mekhilef, M. Seyedmahmoudian, A. Stojcevski, and B. Horan, "Optimized support vector regression-based model for solar power generation forecasting on the basis of online weather reports," *IEEE Access*, vol. 10, pp. 15594-15604, 2022.
- [9] C. Persson, P. Bacher, T. Shiga, and H. Madsen, "Multi-site solar power forecasting using gradient boosted regression trees," *Solar Energy*, vol. 150, pp. 423-436, 2017.
- [10] Sheng, H., Xiao, J., Cheng, Y., Ni, Q. and Wang, S., "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 300-308, 2017.
- [11] Tang, Y., Yang, K., Zhang, S. and Zhang, Z., "Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy," *Renewable and Sustainable Energy Reviews*, vol. 162, p. 112473, 2022.
- [12] Li, Z. and Chen, Z., "Short-term load forecasting based on CEEMDAN-FE-ISSA-LightGBM model," *Frontiers in Energy Research*, vol. 11, p. 1111786, 2023.
- [13] Yu, Z., Ma, J., Qu, Y., Pan, L. and Wan, S., "PM2.5 extended-range forecast based on MJO and S2S using LightGBM," *Science of The Total Environment*, vol. 880, p. 163358, 2023.
- [14] Peng, Y., Wang, S., Chen, W., Ma, J., Wang, C. and Chen, J., "LightGBM-Integrated PV Power Prediction Based on Multi-Resolution Similarity," *Processes*, vol. 11(4), p.1141, 2023.
- [15] N. Rahimi, S. Park, W. Choi, B. Oh, S. Kim, Y. H. Cho, S. Ahn, C. Chong, D. Kim, C. Jin, and D. Lee, "A comprehensive review on ensemble solar power forecasting algorithms," *Journal of Electrical Engineering & Technology*, 2023.
- [16] G. L. Ke, Q. Meng, T. Finley, T. F. Wang, W. Chen, W. D. Ma, Q. W. Ye, and T. Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [18] Friedman, J.H., "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp.1189-1232, 2001.
- [19] Copernicus climate change service, "ECMWF Reanalysis v5 (ERA5)," European Centre for Medium-Range Weather Forecasts <https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5> (accessed 02,15, 2023).
- [20] T. Q. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.
- [21] Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M., "Optuna: A next-generation hyperparameter optimization framework," *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623-2631, 2019.

Curriculum Vitae

Personal data

Name: Hossein Nourollahi Hokmabad
Date of birth: 16-09-1989
Place of birth: Tabriz, Iran
Citizenship: Iranian

Contact data

E-mail: Hossein.Nourollahi@taltech.ee

Education

2022–2025 PhD, Tallinn University of Technology, Estonia
2012–2014 MSC, Tabriz University, Iran
2007–2011 BSC, Tabriz University, Iran
2003–2007 High school, Shahid Madani, Iran

Language competence

English Fluent
Persian Native
Azeri Native
Turkish Fluent
German Basic

Professional employment

2025 School of Information Technologies, Department of Software Science, Research Assistant (0,50)

Elulookirjeldus

Isikuandmed

Nimi: Hossein Nourollahi Hokmabad
Sünniaeg: 16-09-1989
Sünnikoht: Tabriz, Iran
Kodakondsus: Iranian

Kontaktandmed

E-post: Hossein Nourollahi Hokmabad

Hariduskäik

2022–2025 PhD, Tallinn University of Technology, Estonia
2012–2014 MSC, Tabriz University, Iran
2007–2011 BSC, Tabriz University, Iran
2003–2007 Keskkharidus, Shahid Madani, Iran

Keelteoskus

English Edasijõudnud
Persian Kõrgtase
Azeri Kõrgtase
Turkish Edasijõudnud
German Algtase

Teenistuskäik

2025 Tallinna Tehnikaülikool, Infotehnoloogia teaduskond,
Tarkvarateaduse instituut, teadusassistent (0,50)

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-416-2 (PDF)