Engineering Applications of Artificial Intelligence 159 (2025) 111650

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Deep reinforcement learning approach for hybrid renewable energy systems optimization

Inoussa Legrene [a,*], Tony Wong [a], Louis-A. Dessaint [b]

[a] Systems Engineering Department, École de Technologie Supérieure, 1100 Notre Dame West Street, Montréal, Québec, Canada, H3C 1K3
[b] Electrical Engineering Department, École de Technologie Supérieure, 1100 Notre Dame West Street, Montréal, Québec, Canada, H3C 1K3

A R T I C L E I N F O

Keywords:
Batteries
Deep reinforcement learning
Generators
Hybrid renewable energy
Sizing
Solar panels
Wind turbines

A B S T R A C T

The sizing of hybrid renewable energy systems (HRES) is a major challenge faced in contemporary energy research. The optimal configuration based on the specific consumption requirements is essential for strategic energy planning. Effective sizing must balance the investment costs, reliability, environmental impacts, and greenhouse gas emissions while satisfying the expected energy requirements. This study proposes a novel multi-criteria sizing approach based on deep reinforcement learning (DRL). The DRL agent is guided by a reward function that integrates three essential performance metrics: energy cost (LCOE), renewable energy fraction (REF), and the loss of power supply probability (LPSP). A penalty function is also included to consider the reliance on external sources, such as diesel generators and the public grid, promoting greater autonomy and renewable usage. The DRL-based approach was implemented and tested on three distinct demand profiles, using hourly data for one year. A comparative analysis was conducted against three established methods: particle swarm optimization (PSO), multi-objective PSO (MOPSO), and non-dominated sorted genetic algorithm (NSGA-II). The results indicate that DRL significantly outperforms all the benchmark methods in terms of economic efficiency. DRL achieves a significant reduction in the energy costs, ranging from 21.33 % to 30.09 % when compared with PSO, 27.89 %–30.27 % when compared with MOPSO, and 27.63 %–28.47 % when compared with NSGA-II. These findings demonstrate that DRL presents a robust and adaptive framework for the sizing and operational control of HRES. DRL presents more autonomous, cost-effective, and scalable renewable energy solutions by minimizing the energy costs while maintaining the system reliability.

Acronyms and Nomenclature

Acronyms	
AC	Alternating Current
A3C	Asynchronous Advantage Actor-Critic
BESS	Battery Energy Storage System
BG	Biogas Generator
CO_2	Greenhouse gas emission
CRF	Recovery Factor
DC	Direct Current
DG	Diesel Generator
DoD	Depth of Discharge
DQN	Deep Q Network
DRL	Deep Reinforcement Learning
EFCS	Electrolyzer Fuel Cell Energy Storage System
GA	Genetic Algorithm
GEN	Generator
GWO	Grey Wolf Optimizer

(continued on next column)

(*continued*)

Acronyms	
HOMER	Hybrid Optimization Model for Energy Resources
HRES	Hybrid Renewable Energy System
IGWO	Improved Grey Wolf Optimizer
LCOE	Levelized Cost of Energy
LPSP	Loss of Power Supply Probability
MOPSO	Multi Objective Particle Swarm Optimization
MPA	Marine Predator Algorithm
NOCT	Nominal Operating Temperature of Photovoltaic Cells
NPC	Net Present Cost
NREL	National Renewable Energy Laboratory
NSGA-II	Non-dominated Sorted Genetic Algorithm
OSSO	Opposite Social Spider Optimization
PV	Photovoltaic
PSO	Particle Swarm Optimization
REF	Renewable Energy Fraction

(continued on next page)

* Corresponding author.
E-mail addresses: inoussa.legrene.1@ens.etsmtl.ca (I. Legrene), tony.wong@etsmtl.ca (T. Wong), louis-A.dessaint@etsmtl.ca (L.-A. Dessaint).

https://doi.org/10.1016/j.engappai.2025.111650
Received 11 March 2025; Received in revised form 19 May 2025; Accepted 27 June 2025
Available online 16 July 2025

(continued)

Acronyms	
SOC	State of Charge
SSA	Salp Swarm Algorithm
SSO	Social Spider Optimization
TRPO	Trust Region Policy Optimization
WT	Wind Turbine
Symbols	
E_a	Energy bought from grid
E_v	Energy sells on the grid
E_{ch}	Load Energy
E_{dis}	Energy available in the BESS
h_{rf}	Reference height
$I(t)$	Solar irradiance at time t
I_{nom}	Solar irradiance at standard operating conditions
N_{PV}	Number of PV panels
N_{WT}	Number of wind turbine
P_{DG}	Power Generation by Generator
$P_{rated\,DG}$	Diesel generator rated power
P_{Grid}	Power from Connected Grid
P_L	Load demand
P_{PV}	Power generation by PV
$P_{rated\,PV}$	PV panel rated power
P_{WT}	Power generation by wind turbine
$P_{rated\,WT}$	Wind turbine rated power
S_{PV}	PV installation area
T_a	Ambient temperature
$T_c(t)$	PV panel temperature at the time t
T_{nom}	Temperature at standard operating conditions
$v(t)$	Speed at the wind turbine hub at the time t
$v_{rf}(t)$	Speed at the reference height
Greek symbols	
η_{cab}	Efficiency of the wiring
η_{ch}	BESS charge efficiency
η_{dis}	BESS discharge efficiency
$\eta_{conv_{PV}}$	Inverter efficiency on PV side
$\eta_{conv_{BESS}}$	Inverter efficiency on BESS side
η_{PV}	PV efficiency
η_r	PV effectiveness
k_p	Maximum power coefficient of temperature
τ_a	Rate of purchase energy from grid
τ_v	Rate of sale of energy on the grid
α	Coefficient of friction
γ	Rate of energy permissible via the grid
τ	Rate for energy permissible by the generator

1. Introduction

Hybrid renewable energy systems (HRES) are being increasingly implemented owing to the need to reduce greenhouse gas (CO_2) emissions and the diversification of energy sources. However, the process of sizing a hybrid energy system remains a complex challenge owing to the variability of renewable sources and the specific requirements of each consumer. Consequently, several approaches have been developed, including the use of metaheuristic methods such as particle swarm optimization (PSO) (Faria et al., 2023; Musa and Ibrahim, 2015), genetic algorithm (GA) (Medghalchi and Taylan, 2023/10), and grey wolf optimization (GWO) (Mahmoud et al., 2022/11), along with the use of simulation tools such as HOMER/HOMER Pro (Bazzi et al.; Lilienthal, 2005). These methods present excellent performance in the various contexts in which they have been used. However, they face several limitations, which are given below:

- Sensitivity to the initial optimization parameters—metaheuristics require precise, algorithm-specific parameter adjustments to avoid being trapped in the local optima.
- Partial consideration of dynamic variations—variation in the loads and generation sources must be considered to ensure a system that is robust against the intermittent nature of the generation sources.
- High computational cost—some approaches, such as dynamic programming and simulation tools, require a considerable amount of time and resources to optimize large ensembles.

Therefore, in this study, we proposed a method based on a deep reinforcement learning (DRL) approach.

The application of DRL presents significant advantages in terms of the adaptability to varying conditions, thereby creating a balance between exploration and exploitation, multi-criteria management, and flexibility for different configurations. By applying Deep RL, this study promotes four key aspects: adaptability and continuous learning, a balance between exploration and exploitation, flexibility and generalization, and reward-based learning.

Since HRES face constant variations in the production and demand, DRL enables real-time adaptation to these variations, thereby improving the system efficiency by continuously adjusting the energy management strategies. Exploration and exploitation must be balanced to achieve better interplay between energy consumption and maximizing the long-term efficiency. The proposed approach can be applied to various HRES to integrate new technologies (thermal storage, gas generators) or facilitate adaptation at different scales (from microgrids to large-scale grids). Reward-based learning helps in achieving the perfect compromise between multiple objectives, which is crucial since different and often conflicting objectives (system cost, energy efficiency, emission reduction) must be considered for the sizing of a reliable HRES. Reinforcement learning makes it possible to model these trade-offs using a reward function, enabling an efficient multi-criteria optimization that considers the variability of renewable energy sources.

Based on these observations, this study addresses four issues that have not yet been addressed in the previous studies conducted on HRES. (i) Adaptive sizing: component capacities (PV, WT, BESS, GEN) are learned online using a deep reinforcement learning policy, addressing the conventional "sizing-dispatch" dichotomy. (ii) Multi-criteria cumulative reward: a unified reward signal simultaneously reduces the levelized cost of energy (LCOE) and loss of power supply probability (LPSP), while maximizing the fraction of renewable energy (REF), without requiring the prior heuristic weighting or a posteriori Pareto front sorting. (iii) Flexible simulator-agnostic framework: the algorithm interacts with the environment only through state-action pairs, enabling a high-fidelity digital twin or a data-driven metamodel to be freely interchanged. (iv) Direct integration of generator-grid constraints: generator ramp limits and grid purchase/sale quotas are encoded in the Markov decision process, thereby ensuring operational feasibility in the island, grid-coupled, or hybrid modes. These four factors constitute the novelty of this study and pave the way for real-time, multi-criteria, and scalable optimization of HRES.

The main contributions of this study can be summarized as follows:

- **Adaptive and Autonomous Sizing of Renewable Energy Systems Using Deep Reinforcement Learning (DRL)**—Unlike conventional metaheuristic approaches that require manual parameter tuning, DRL learns autonomously from system interactions, making it more adaptable to variations in the energy demand and renewable resource availability. This extends beyond fixed-rule optimization by enabling self-learning strategies that improve with time.
- **Multi-Criteria Cumulative Reward Modeling for Hybrid Renewable Energy Systems Optimization**—One of the major challenges in HRES sizing involves balancing conflicting objectives, such as cost minimization, renewable energy utilization, and system reliability. This study presents a reward-based framework that dynamically evaluates the trade-offs between these factors, enabling the DRL model to optimize the energy dispatch and storage decisions in real time while ensuring a cost-effective and reliable system configuration.
- **Flexible and Scalable Optimization Framework for Hybrid Renewable Energy Systems**—The proposed DRL-based methodology is designed to be adaptable to various HRES architectures, from off-grid microgrids to large-scale grid-connected systems.

The remainder of this article is organized as follows. Section 2 presents an overview of recent studies on the design of hybrid renewable energy systems, including their methodologies and types of systems. Section 3 presents the modeling of the HRES, the problem formulation, and the proposed systematic DRL approach. Section 4 presents the results, discusses the methods developed in this study, and compares the results with those reported in previous studies. Lastly, Section 5 presents the conclusion and future research implications.

2. Related works

Extensive research has been conducted on HRES. However, there is no unanimity on the components (photovoltaic (PV) solar panels, wind turbines (WT), batteries as an energy storage system (BESS)) that constitute these hybrid systems (Faria et al., 2023; Kushwaha and Bhattacharjee, 2024; Samy et al., 2018), the best algorithm or optimization tool (Musa and Ibrahim, 2015; Baidas et al., 2022), or the cost, reliability, and feasibility evaluation factors required for system performance analysis (Kushwaha and Bhattacharjee, 2024; Kushwaha et al., 2022). The diversity of these systems can be attributed to the uniqueness of each problem, which is characterized by a proprietary consumption profile in various weather conditions that affect the availability of the exploitable energy sources.

Kushwaha and Bhattacharjee (2024) reported various systems based on PV solar panels, WT, BESS, and biogas (BG) and diesel (DG) generators as an external compensation system for the difference in production. Following several evaluations, they obtained an optimal PV-WT-BESS-BG-DG system based on social, economic, technological, and environmental criteria, with an energy cost of 0.1799 $/kWh for a village in India. Mokhtara et al. (Mokhtara et al., 2021/03) also proposed an optimal design approach for PV/WT/BESS/DG hybrid systems for the electrification of residential buildings in rural areas, considering the energy performance of the buildings and the climatic conditions. Medghalchi and Taylan (Medghalchi and Taylan, 2023/10) addressed the sizing of a system comprising PV, WT, BESS, and a fuel cell electrolyzer (Lu et al., 2017). They reported that a combination of PV, WT, and BESS is the most favorable system, with an energy cost of 0.1838 Euro/kWh. Bouafia et al. (Bouafia et al., 2023) evaluated various combinations comprising PV and WT in 12 localities in Morocco. They reported that the best system is obtained by integrating PV and WT. Lastly, several studies, including the analyses conducted by Samy et al. (Samy et al., 2021), demonstrated that green energy systems (PV/WT) can improve the reliability of unstable power grids while highlighting the economic implications.

Various methods have been developed in the previous studies. Simulation tools such as HOMER/HOMER Pro (Bazzi et al.; Lilienthal, 2005) and PVGIS (Lemaire et al.) have been used to simulate the typical examples of predefined systems (configurations). Several works have highlighted the economic and feasibility factors of this simulation approach as the criteria for selecting the optimal system (Baidas et al., 2022; Bahgaat, 2023/10; Lu et al., 2017). Baidas et al. (Sandeep and Nandihalli, 2020) simulated various configurations using HOMER, which were then evaluated and compared. The comparison criterion was the net present cost (NPC), with the most profitable configuration from an economic and environmental perspective involving integrating WTs and batteries. Similarly, several studies have adopted metaheuristic algorithms, such as PSO and its derivatives, including MOPSO, instead of simulation tools (Faria et al., 2023; Musa and Ibrahim, 2015; Medghalchi and Taylan, 2023/10; Bouafia et al., 2023; Samy et al., 2021; Mansouri Kouhestani et al., 2020; Ukoima et al.; Samy and Barakat, 2019), GA (Medghalchi and Taylan, 2023/10) and its derivatives, such as the non-dominated sorted genetic algorithm II (NSGA-II) (Ukoima et al.; Sandeep and Nandihalli, 2020), Salp Swarm Algorithm (SSA), GWO and its derivatives (Mahmoud et al., 2022/11), and social spider optimization (SSO) and opposite social spider optimization (OSSO) (Sandeep and Nandihalli, 2020). Faria et al. (2023) analyzed MOPSO to analyze the shared energy management between several consumers, also known as community management. Economic and technical criteria, such as the energy costs and self-sufficiency ratios, were in direct conflict in this work, and based on the analytical approach adopted, the so-called optimal configuration varies between the members of the community and between the approaches used to exploit the energy produced (Lemaire et al.). In Kushwaha and Bhattacharjee (2024), several metaheuristic algorithms are used to determine the best approach, including the marine predator algorithm (MPA), GA, PSO, and SSA. The MPA is significant due to its ability to evade local minimums while obtaining the best energy cost, at 0.1799 $/kWh, for the best configuration. Similarly, Mahmoud et al. (Mahmoud et al., 2022/11) compared SSA, GWO, and improved GWO (IGWO). In they study, they primarily focused on energy production using PV panels and WTs, BESS, and external energy production using a diesel generator. Consequently, they demonstrated that the IGWO method presented the best performance when compared with the other approaches.

A wide range of factors can be considered as the selection criteria. Typically, various factors, such as the algorithm or simulation tool used, performance evaluation criteria, components of the HRES, weather conditions affecting the choice of renewable sources, and energy requirements to be met, significantly affect the final system to be adopted. Although the metaheuristic methods present good performance when the parameters are well defined, they can also underperform and get trapped in the local optima (Lilienthal, 2005) if the environmental parameters are not defined precisely. Essentially, the solution approach depends on the problem to be solved and requires adjusting the hyperparameters for each problem instance.

3. System and problem formulation

In this study, we addressed the sizing of HRES, as shown in Fig. 1 (Legrene et al., 2025). PV panels and WTs have been used as the main sources of energy production in various research projects (Baidas et al., 2022; Al-Quraan and Al-Mhairat, 2024/07). These renewable sources are supported by generators running on natural gas or diesel (GEN). In particular, generators are included to ensure the system reliability, particularly in the event of extreme weather conditions or grid outages. This study minimizes their usage to reduce the carbon footprint and maximize the use of renewable energies. Additionally, these systems are typically combined with BESS to make them robust (Mahmoud et al., 2022/11; Baidas et al., 2022). Consequently, we considered four different types of HRES (based on the principle depicted in Fig. 1):

- PV solar panels coupled with BESS: PV + BESS;
- PV solar panels coupled with BESS and generators: PV + BESS + GEN;
- PV solar panels and WTs coupled with BESS: PV + WT + BESS; and
- PV solar panels and WTs coupled with BESS and generators: PV + WT + BESS + GEN.

The following subsection present the mathematical approaches applied to the system elements (PV solar panels, WTs, BESS, GEN, and hybrid converter), the problem formulation, and the optimization strategy.

3.1. System modeling

3.1.1. Solar panel model

Several factors must be considered when producing energy through PV systems, including solar irradiance, ambient temperature, and panel efficiency (Mahmoud et al., 2022/11; Tabak et al., 2017; Memon et al., 2021). The energy produced depends on the number of panels in the system, which is an optimization problem parameter. An optimization method is implemented to determine the minimum number of PVs that help achieve the ideal compromise between the amount of energy produced and the cost. The amount of energy produced is defined as

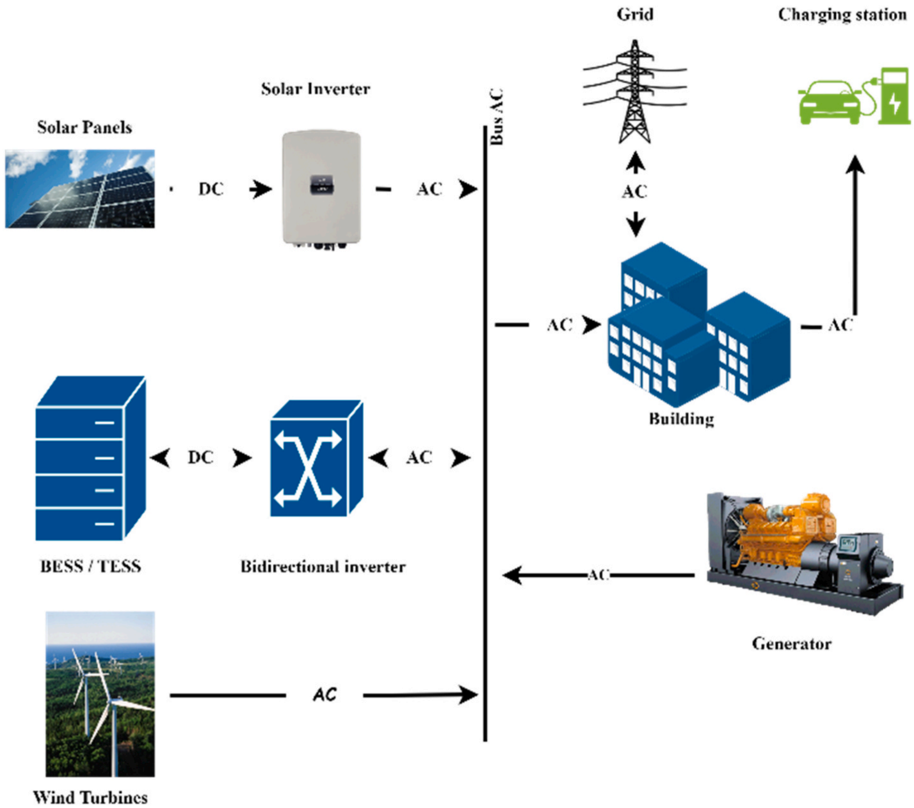

Fig. 1. Hybrid renewable energy system (HRES) (Legrene et al., 2025).

follows:

$$P_{\mathrm{PV}} = \eta_{\mathrm{PV}} \times S_{\mathrm{PV}} \times I(t) \quad (1)$$

where η_{PV} and S_{PV} denote the instantaneous efficiency and the installation area of the panels (in m^2), respectively, and are defined as follows (Mahmoud et al., 2022/11; Memon et al., 2021):

$$\eta_{\mathrm{PV}} = \eta_r \times \left(1 - \beta_T (T_c(t) - T_{\mathrm{nom}})\beta_T = \frac{k_p}{100} T_c(t) = T_a + \left[\left(\frac{\mathrm{NOCT} - 20}{800}\right) \times I(t)\right]\right. \quad (2)$$

$$S_{\mathrm{PV}} = \frac{N_{\mathrm{PV}} \times P_{\mathrm{rated_{PV}}} \times \eta_{\mathrm{cab}}}{I_{\mathrm{nom}}} \quad (3)$$

Here, η_r denotes the effectiveness of the panel, T_{nom} and I_{nom} denote the standard operating conditions ($T_{\mathrm{nom}} = 25\ ^\circ\mathrm{C}$ and $I_{\mathrm{nom}} = 1000\ \mathrm{W/\ m^2}$), T_a and T_c denotes the ambient temperature (based on the locality) and the temperature, respectively, of the panel (in $^\circ$ C). Additionally, kp denotes the maximum power coefficient of temperature (in %/ °C), and NOCT denotes the nominal operating temperature of the PV cells, which is provided by the manufacturer. $I(t)$ denotes the solar irradiance at time t (in $\mathrm{W/m^2}$), $P_{\mathrm{rated_{PV}}}$ denotes the nominal output power of the panels, and η_{cab} denotes the efficiency of the wiring, which are considered as the input data of the optimization problem. N_{PV} denotes the number of solar panels that need to be installed to satisfy the load demand, the principal solar parameter of the optimization problem. Lastly, P_{PV} (in W) denotes the amount of energy produced by the installed solar system. In this study, we considered monofacial solar panels, with the parameters defined in Table 1 (Faria et al., 2023; "18 KW All in 1 V).

3.1.2. Wind turbine model

The energy production process using WTs involves converting the kinetic energy (due to the speed) of the wind into electrical energy. The energy production of a wind turbine installation is expressed as follows (Tabak et al., 2017; Memon et al., 2021):

$$P_{\mathrm{WT}} = N_{\mathrm{WT}} \times \eta_{\mathrm{WT}} \times \begin{cases} 0\ si\ \ v(t) < \ v_{ci}, v(t) > \ v_{co} \\ P_{\mathrm{rated_{WT}}} \times \dfrac{v^3 - v_{ci}^3}{v_r^3 - v_{ci}^3}, v_{ci} < v(t) < v_r \\ P_r, v_r \le v(t) \le v_{co} \end{cases} \quad (4)$$

where N_{WT}, η_{WT}, and $P_{\mathrm{rated_{WT}}}$ denote the number of WTs, turbine efficiency, and rated power output, respectively, of each turbine (in kW or MW). These parameters are typically provided by the manufacturer. The variables, v_{ci}, v_{co}, and v_r denote the wind cut-off speed (in m/s) at which the turbine produces power, the cut-off speed at which the turbine is stopped, preventing damage, and thus ceases to produce energy, and the rated speed at which the turbine produces energy equivalent to its rated power output, respectively. The v_{ci} and v_{co} values are typically provided by the manufacturer. v denotes the speed of the WT at the hub and is defined as follows (Memon et al., 2021; Al-Ghussain et al., 2020/04):

Table 1
PV specifications.

Parameters	Symbols	Value	Unit
Nominal Output	$P_{\mathrm{rated_{PV}}}$	200	W
Cell efficiency	η_r	22.90	%
NOCT	NOCT	45	°C
PV Cost	–	367	$/unit
Installation cost	–	1.27	$/W
Maintenance and Op. Cost	–	0.015	$/W
CO_2 Emission	–	0.0338	$kgCO_2$/kWh
Irradiance at SC	–	1000	W/m2
Temperature at SC	–	25	°C
Temp. Coef of Voc	kp	−0.29	%/°C
PV Lifetime	–	25	years
Cabling efficiency	η_{cab}	95	%

$$v(t) = v_{rf}(t) \times \left(\frac{h_{rf}}{\mathrm{h_{hub}}}\right)^{\propto} \quad (5)$$

where v_{rf}, h_{rf}, and h_{hub} denote the speed at the reference height (in m/ s), reference height at which the data are available, and height of the WT hub, respectively, and $\propto$ denotes the coefficient of friction. The BWC XL-1 model was used for the experiments; Table 2 presents its technical description (Faria et al., 2023; Al-Quraan and Al-Mhairat, 2024/07).

3.1.3. Battery energy storage system

The batteries in HRES conserve the surplus produced from renewable sources (solar, wind) when all the energy that is produced is not consumed by the loads. This energy, which is stored during over-production, is used when the energy demand is greater than the energy produced by the system. Defining the battery capacity size to be used in HRES depends on several factors, including the battery life, acceptable battery discharge threshold, and ambient temperature (Mahmoud et al., 2022/11; Ismail et al., 2013). The state of charge (SOC), or the ratio of stored energy to battery capacity, is an essential parameter during the charging and discharging of batteries (Mahmoud et al., 2022/11).

3.1.3.1. BESS charging process. Battery charging in HRES occurs when the energy production by renewable sources is greater than the demand at a given time, *t*. The battery charge is defined as follows (Al-Quraan and Al-Mhairat, 2024/07; Memon et al., 2021):

$$E_{\mathrm{ch}} = \frac{\left[\left(P_{\mathrm{WT}}(t) + \eta_{\mathrm{conv_{PV}}} \times P_{PV}(t)\right) - P_L(t)\right]}{\eta_{\mathrm{conv_{BESS}}}} \times \eta_{\mathrm{ch}} \quad (6)$$

$$\mathrm{SOC}(t) = SOC(t-1)(1-\delta) + E_{\mathrm{ch}}(t) \quad (7)$$

where $\eta_{\mathrm{conv_{PV}}}$, $\eta_{\mathrm{conv_{BESS}}}$, and η_{ch} denote the efficiency of the solar converter and battery and charging efficiency of the battery system, respectively. $P_L(t)$ and E_{ch} are the load demand and the load energy at time *t*, and SOC(t) and SOC$(t-1)$ in Eq. (7) represent the battery's state of charge at times *t* and $t-1$, respectively.

3.1.3.2. BESS discharge process. The battery discharge process must be automatic when the total energy produced by renewable sources in the system (in this case, PV and WT) cannot sufficiently meet the load demand. The battery discharge must consider the allowable depth of discharge (DoD) and the self-discharge coefficient. The discharge is expressed as follows (Al-Quraan and Al-Mhairat, 2024/07; Memon et al., 2021):

$$E_{\mathrm{dis}} = \frac{\left[P_L(t) - \left(P_{\mathrm{WT}}(t) + \eta_{\mathrm{conv_{PV}}} \times P_{PV}(t)\right)\right]}{\eta_{\mathrm{conv\_BATT}}} \times \eta_{\mathrm{dis}} \quad (8)$$

$$\mathrm{SOC}(t) = \mathrm{SOC}(t-1)(1-\delta) - E_{dis}(t)$$

where η_{dis} denotes the discharge efficiency of the battery (accounting for the accepted discharge threshold), and δ denotes the self-discharge coefficient. $E_{\mathrm{dis}}(t)$ denotes the energy available in the battery that can be applied to the power loads.

3.1.3.3. BESS specifications. Table 3 presents the specifications of the Elios EA12-200 battery model, which was used in this study ("18 KW All in 1 V).

3.1.4. Diesel generators

In this context, energy was produced using diesel-powered generators. A generator is used only when the energy produced by renewable sources added to the energy available in the storage sources cannot sufficiently meet the load demand. However, this study, which also focuses on limiting the greenhouse gas emissions, defines a production limit for the generator (e.g., it is not allowed to produce more than 20 % of the total load demand).

The energy output of a diesel generator corresponds to the amount of fuel consumed, C_{DG}, which is defined as follows (Mahmoud et al., 2022/11; Ismail et al., 2013):

$$C_{DG} = \alpha_{\mathrm{DG}} \times P_{\mathrm{DG}}(t) + \theta_{\mathrm{DG}} \times P_{\mathrm{rated_{DG}}} \quad (9)$$

where $P_{\mathrm{DG}}(t)$ and $P_{\mathrm{rated_{DG}}}$ denote the average and the nominal power of the generator, respectively. The variables, α_{DG} and θ_{DG}, denote the coefficients of the consumption curve and are equal to 0.246 and 0.08145 l/kWh, respectively. We considered the Generac generators in the context of these experiments; Table 4 presents the specifications of these generators (Mahmoud et al., 2022/11).

3.1.5. Hybrid converter

The HRES considered in this study requires the use of solar converters, which can convert the electricity produced by the PV panels in direct current (DC) into alternating current (AC), suitable for consumption by devices and injectable into the electricity grid, while protecting against overvoltage incidents and short circuits. On the energy storage side, a bidirectional converter converts the AC electricity from the system into DC to recharge the batteries. It converts the DC energy stored in the batteries into AC to be consumed by the loads. We used the 12 kW ELIOS hybrid converter that is suitable for both PV and BESS. Table 5 presents its specifications ("18 KW All in 1 V).

3.2. Energy management strategies

Different scenarios can affect the economic factors, feasibility, and reliability of HRES (Faria et al., 2023). Fig. 2 depicts the process of managing HRES.

In this process, we considered two renewable sources of energy production (solar and wind), an energy storage system (lithium battery), and a diesel generator (Fig. 1).

At each given time, *t*, the total energy produced by the system (solar and wind) is calculated and compared with the load demand. When production is sufficient, two possible scenarios are observed:

1) When the energy produced is equivalent to the demand: In this case, the load consumes all the energy produced, and no amount of energy

Table 2
WT specifications (Faria et al., 2023).

Parameters	Symbols	Value	Unit
Rated power	P_r	1	kW
Turbine efficiency	η_{WT}	96	%
Cut-in Speed	v_{ci}	2.5	m/s
Cut-out Speed	v_{co}	20	m/s
Rated speed	v_r	11	m/s
Height at WT hub	h_{hub}	10	M
Height at reference	h_{rf}	20	M
Lifetime	–	30	years
Capital cost	–	1800	$/kW
Op. & Maint. Cost	–	15	$/kW/year
Friction coef.	$\propto$	0.35	–

Table 3
BESS specifications.

Parameters	Symbols	Value	Unit
State of Charge min	SOCmin	10	%
State of Charge max	SOCmax	100	%
Charge-Discharge efficiency	η_{ch}, η_{dis}	92	%
Auto discharge rate	δ	0.01	%
Power rating	–	200, 12	Ah, V
Lifetime	–	15	years
Capital Cost	–	495	$/unit
Op. & Maint. Cost	–	5	$/year/unit

Table 4
Generator specifications.

Parameters	Value	Unit
Capacities	7.5, 10, 13, 14, 16, 18, 20, 22, 24	kW
Fuel consumption	(2.07, 3.31), (2.86, 3.60), (4.36, 6.37), (5.52, 7.25), (5.15, 6.94), (4.79, 6.99), (5.78, 8.52), (6.46, 9.26), (5.75, 8.66)	(50 %, 100 %) [m^3/hr]
Operation load	95	%
Capital cost	370	$/kW
Op. & Maint. Cost	3	%
Fuel cost	1.2	$/L
Carbon emission	66	kg/kW
Carbon emissions from fuel	2.6	kg/L
Lifetime	25	Years

Table 5
12 kW ELIOS hybrid converter specifications.

Parameters	Value	Unit
Maximum charging/discharging current	250/250	A
Maximum charging/discharging power	12	kW
Battery charging efficiency	95	%
Battery discharging efficiency	94.50	%
Capital cost	300	$/kW
Op. & Maint. cost	5	$/year
Lifetime	15	year

is imported from the batteries or the grid, nor used to charge the batteries or injected into the grid.

2) When the energy produced is greater than the demand: In this case, the difference in the production and load demand that represents the surplus is initially used to charge the battery until the maximum allowable SOC is reached. When the surplus production is not exhausted after charging the storage system to the maximum allowable, it is fed into the grid at the defined rate, μ. When the storage system is unable to take any amount of surplus, all of it is injected into the grid at the set rate.

Two scenarios are observed when the production cannot meet the load demand at time t.

1) The energy storage system contains sufficient energy to cover the difference: The storage system (the battery) is discharged to satisfy the difference in energy between the demand and production.
2) The storage system does not contain sufficient energy to cover the difference: The generator is used to produce energy, considering its production condition. Here, τ denotes the limit of permissible energy obtained from the generator. This approach presents two forms of management. The first is if the limit imposed on the generator enables the amount of power to cover the demand. The second is if the energy produced by the generator does not cover the difference in the energy demand, considering the imposed limit. In this case, after using the energy stored in the storage system and the amount safely generated by the generator, the remaining demand is imported from the grid at the set rate, ϑ.

The rates, ϑ and μ, are calculated as 0.14 and 0.05 $/kWh, respectively (Medghalchi and Taylan, 2023/10; Bouafia et al., 2023; Al-Ghussain et al., 2020/04).

3.3. System analysis factors

Previous studies considered the net present cost (NPC) (Bahgaat, 2023/10; Mansouri Kouhestani et al., 2020), LCOE (Bouafia et al., 2023; Legrene et al., 2024), and weighted average cost of energy (waCOE) (Medghalchi and Taylan, 2023/10) as factors to evaluate the economics of HRES. Similarly, Baidas and Al-Quraan (Baidas et al., 2022; Al-Quraan and Al-Mhairat, 2024/07) used the REF to analyze the integration of renewable sources in a hybrid system. Since the reliability of the system is equally important, Mansouri Kouhestani et al. (Mansouri Kouhestani et al., 2020) employed LPSP to evaluate the reliability of these systems. Cheraghi and Hossein Jahangir (Cheraghi and Hossein Jahangir, 2023/10) proposed the use of the human development index (HDI) to determine the system reliability. Since different factors can be used to analyze HRES, the sizing process evaluated here considers the LCOE, REF, and LPSP.

3.3.1. Levelized Cost Of Energy—LCOE

The LCOE is the measurement factor used to define the cost (in $/kWh) of the energy produced by a system. This coefficient must be minimized while sizing HRES. The total cost of a system is determined based on the ratio of the HRES LCOE $\mathbb{C}_{\text{Total}}$ (in $) to the total load for a year (in kWh), which is given as follows (Eq. (10)):

$$LCOE = \frac{\mathbb{C}_{\text{Total}}}{\sum P_L} \tag{10}$$

The total cost of a system includes the initial investment capital, $\mathbb{C}_{\text{Capital}}$, costs associated with replacements, $\mathbb{C}_{\text{Remp}}$, and the operations and maintenance costs, $\mathbb{C}_{\text{Op\&Maint}}$ (Coban, 2024/09). Additionally, for HRES connected to the electricity grid, the cost of the energy purchased from the grid and the total cost of sales must also be considered (Coban et al., 2024). The total cost, $\mathbb{C}_{\text{Total}}$, is expressed as follows (Memon et al., 2021; Coban et al., 2024):

$$\mathbb{C}_{\text{Total}} = \mathbb{C}_{\text{Capital}} + \mathbb{C}_{\text{Remp}} + \mathbb{C}_{\text{Op\&Maint}} + E_a \times \tau_a - E_v \times \min(\tau_a, \tau_v) \tag{11}$$

where E_a, E_v , τ_a, and τ_v denote the quantities of energy bought and sold and the rates of purchases and sales on the grid, respectively.

- Capital cost

The total cost of capital involves the sum of the product of the initial cost of each piece of equipment, j, $\mathbb{C}_{\text{Initial}_j}$, multiplied by its recovery factor, $CRF(r, m_j)$. This initial cost is defined as follows:

$$\mathbb{C}_{\text{Capital}} = \sum \mathbb{C}_{\text{Initial}_j} \times CRF(r, m_j) \tag{12}$$

$$CRF(r, m_j) = \frac{r(1+r)^{m_j}}{(1+r)^{m_j} - 1}$$

$$\mathbb{C}_{\text{Initial}_j} = \text{Cap}_j \times \mathbb{C}_{uj}$$

where Cap_j and $\mathbb{C}_{uj}$ denote the capacity of the PV, WT, DG, and BESS components (in W for PV and WT and in Wh for BESS and DG) and the unit cost of each component (in $/kW or $/W for PV and WT and in $/kWh or $/Wh for ESS and DG), respectively, and where r and m_j denote the interest rates (in %) and the life of the component, j.

- Replacement cost

The replacement cost helps in preventing a component from malfunctioning during the lifetime of the system, m_s. This cost is introduced for each component when m_s is greater than the lifetime of the component, m_j ($m_s > m_j$), and can be expressed as follows:

$$\mathbb{C}_{\text{Remp}} = \sum \mathbb{C}_{\text{Initial}_j} \times \frac{(m_s - m_j)}{m_j} \tag{13}$$

- Maintenance and operation cost

The cost of operation and maintenance, $\mathbb{C}_{\text{Op\&Maint}}$, represents the main cost of the operation of HRES (Mahmoud et al., 2022/11). When

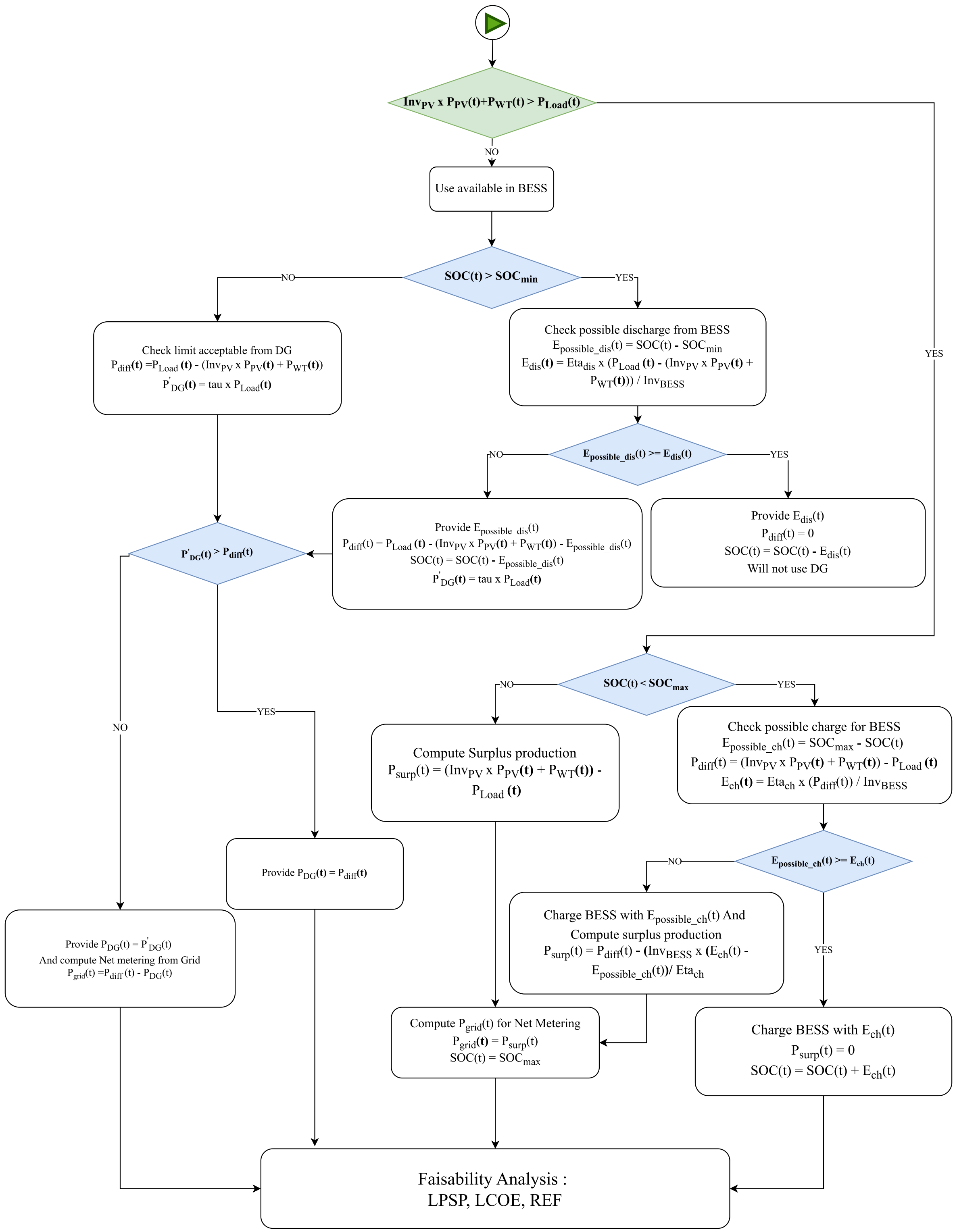

Fig. 2. Energy management process.

this cost cannot be determined annually, it can be calculated based on the product of the operating (or maintenance) cost per hour of each component ($\mathbb{C}_{\mathrm{Op\&Maint}_j}$) and the duration of operation and maintenance ($t_{\mathrm{Op\&Maint}_j}$), as shown below:

$$\mathbb{C}_{\mathrm{Op\&Maint}} = \sum \mathbb{C}_{\mathrm{Op\&Maint}_j} \times t_{\mathrm{Op\&Maint}_j} \tag{14}$$

3.3.2. *Loss of Power Supply Probability - LPSP*

The reliability of HRES is measured based on the value of the LPSP (in %) (Mansouri Kouhestani et al., 2020). This probability helps in quantifying the underproduction of energy by the system (LPS) when compared with the load demand, P_L. This probability is expressed as follows:

$$\mathrm{LPSP} = \frac{\sum_{t=1}^{8760} \mathrm{LPS}(t)}{\sum_{t=1}^{8760} P_L(t)} = \frac{\sum_{t=1}^{8760} (P_L(t) - P_{\mathrm{Total}}(t))}{\sum_{t=1}^{8760} P_L(t)} \tag{15}$$

where $P_L(t)$ and LPS (t) denote the load demand and energy loss at the instant t, respectively.

LPS loss occurs only when all the energy produced (PV + WT) is added to that available in the storage systems, and when these amounts, combined with the additional energy produced by the generator, cannot satisfy the load demand. The acceptable reliability threshold for HRES is 5 % (Mansouri Kouhestani et al., 2020).

3.3.3. Renewable Energy Fraction—REF

The REF is designed to help in selecting HRES whose main production sources are solar and wind. Thus, in HRES that employs a diesel generator, the REF (in %) is defined as follows (Baidas et al., 2022):

$$\mathrm{REF} = \frac{\sum P_{\mathrm{renewable}}}{\sum P_{\mathrm{Total}}} = \frac{\sum (P_{\mathrm{WT}} + \eta_{\mathrm{PV}} \times P_{\mathrm{PV}})}{\sum (P_{\mathrm{WT}} + \eta_{\mathrm{PV}} \times P_{\mathrm{PV}} + P_{\mathrm{grid}} + P_{\mathrm{DG}})} \tag{16}$$

where P_{WT} and $\eta_{\mathrm{PV}} \times P_{\mathrm{PV}}$ denote the energy production from the wind and solar sources, respectively. The variables, P_{grid} and P_{DG}, represent the quantities of energy taken from the electricity grid and produced by the diesel generator, respectively. In a hybrid renewable energy system, the fraction of renewable energy must be maximized, thereby reducing the greenhouse gas emissions.

3.3.4. Constraints

The feasibility constraints affect the reliability criterion, the energy produced by the diesel generator, and the amount of energy taken from the grid. These constraints are summarized in Eq. 17.a, 17.b, and 17.c, respectively. The impact of these constraints on the search for the optimal system is determined based on a set of values considered within the framework of the study.

We set $\varepsilon = 0.05$ (Mansouri Kouhestani et al., 2020), conducted different experiments with γ and τ values of ([0.1, 0.2, 0.3, 0.4]), and proposed a methodical evaluation of the parameters that can affect the performance of the HRES. The selection of the values helps in isolating the effects of the different parameters and in identifying the sensitivities that enable more efficient optimization of the system, owing to a good understanding of the impact of each constraint.

$$\mathrm{LPSP} \leq \mathrm{LPSP}^{max} = \varepsilon \tag{17.a}$$

$$P_{DG} \leq P_{DG}^{max} = \tau \times P_L \tag{17.b}$$

$$P_{\mathrm{grid}} \leq P_{\mathrm{grid}}^{max} = \gamma \times P_L \tag{17.c}$$

$$\varepsilon = 0.05, \gamma \, \epsilon \, [0.1, 0.2, 0.3, 0.4], \tau \, \epsilon \, [0.1, 0.2, 0.3, 0.4]$$

3.4. Problem formulation

The optimization of a hybrid renewable energy system in this framework is a multi-criteria optimization problem under the previously defined constraints, as shown below:

$$\begin{cases} \min f_1 = min\ LOCE \\ \min f_2 = min\ LPSP \\ \max f_3 = max\ REF \end{cases} \tag{18}$$

In the framework of the proposed method, we proposed a DRL optimization approach in which a reward function is defined based on the above objective functions. This reward function in the sizing process is expressed as follows:

$$\mathrm{R} = \sum_{i=0}^{n} \left[\sum_{j=1}^{k} \frac{\pm \left(f_j^{best} - f_j^{i} \right)}{\max\left(\left\{ \left| f_j^{best} - f_j^{i} \right|, j = 1, 2, \dots, k \right\} \right)} + \phi \right] \tag{19}$$

$$\begin{cases} f_j^{best} = f_j^{(i)}, if\ i = 0 \\ f_j^{best} = f_j^{(i)}, if\ f_j^{best} \leq f_j^{(i)} \\ f_j^{best} = f_j^{best}, else \end{cases}$$

where n and k denote the number of iterations and number of objective functions, respectively. Normalizing the differences between new and best values by maximizing them helps avoid prioritizing a single goal function. Additionally, ϕ denotes a cost function corresponding to the search constraints, which is defined as follows:

$$\phi = \frac{P_{sys}^{max} - P_{sys}^{(i)}}{\max\left(\left\{ \left| P_{sys}^{max} - P_{sys}^{(i)} \right| \right\},\ P_{sys} \epsilon \left\{ P_{grid}, P_{DG} \right\} \right)} + \left(LPSP^{max} - LPSP^{(i)} \right) \tag{20}$$

Applying ϕ enables the adjustment of the objectives. Normalization between −1 and 1 helps in avoiding maximization or minimization at the expense of other objectives. Thus, the proposed method aims to maximize the cumulative reward, R.

3.5. Research approach

This section presents the proposed approach and the methods used as a benchmark. Furthermore, we define the framework, including the case studies considered and the values of the parameters.

3.5.1. Deep reinforcement learning

We used DRL with greedy epsilon to optimize the HRES.

Reinforcement learning is a major breakthrough in the fields of machine learning and artificial intelligence for the development of intelligent, autonomous, and emerging technology (Du and Ding, 2021/06; Le et al., 2022/04; Mnih et al., 2015/02; Nguyen et al., 2020). DRL helps in overcoming the inability of conventional reinforcement learning methods to obtain a better understanding of environments with high variability and adapt to complex state systems owing to the usage of neural network models in the reinforcement learning process (Zhu et al., 2023). Advances in the use of deep reinforcement learning models have helped in solving previously unsolvable problems, such as learning to play video games from pixel inputs (Mnih et al., 2015/02; Arulkumaran et al., 2017).

Various DRL models have been developed for solving various problems. Three of these ar: deep Q-network (DQN), which approximates the value of the Q-value function, enabling agents to learn policies from inputs (Mnih et al., 2015/02); trust region policy optimization (TRPO) (Arulkumaran et al., 2017), which ensures that policy optimization is achieved stably and efficiently without deviating significantly from the current policy (Kiran et al., 2022); and asynchronous advantage actor-critic (A3C), which involves parallelizing several agents, promoting stability, and accelerating the learning process that enables it to adapt to complex environments (Kiran et al., 2022).

These models have been analyzed in other areas, such as automatic driving (Kiran et al., 2022), in multi-agent systems, enabling better communication and collaboration in complex tasks, such as non-stationarity and partial observability problems (Nguyen et al., 2020; Kiran et al., 2022), and in neuroscience (Botvinick et al., 2020/08).

3.5.2. Approach

Fig. 3 depicts the reinforcement learning approach adopted in this study, which is based on TD3 with the epsilon-greedy approach. This

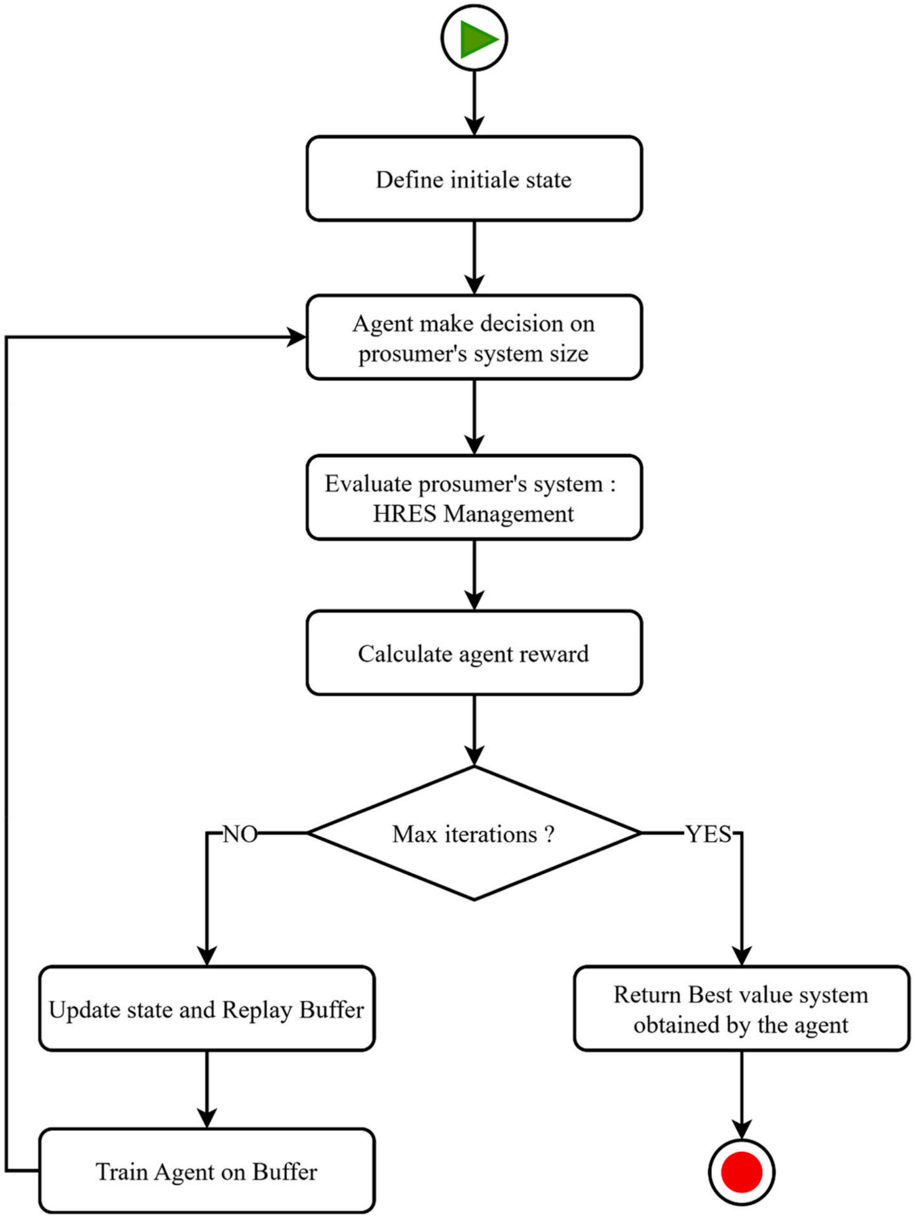

Fig. 3. Proposal approach steps.

approach comprises five main steps: state initialization, agent decision-making, HRES evaluation, agent reward calculation, and agent management and training on the buffer.

3.5.2.1. Initial state. Defining the initial state helps in assigning a state to the system, i.e., a set of values that enables the system to be defined at time 0 of the process. The system state settings always remain unchanged. The values are only updated during the training process, enabling the changes caused by a particular decision of the agent to be known at any given time. The state of the system, S, is then defined as follows:

$$\mathrm{S} = \{N_{PV}, N_{WT}, \mathrm{Cap}_{BESS}, \mathrm{Cap}_{DG}, E_{np}, E_l, C_{LCOE}, C_{LPSP}, C_{REF}\} \tag{21}$$

where N_{PV} and N_{WT} denote the number of photovoltaic solar panels and wind turbines, respectively. Cap_{BESS} and Cap_{DG} denote the capacity of the storage system (battery) and the diesel generator. E_{np} and E_l denote the amount of energy not supplied and the amount of energy lost over the entire evaluation period, respectively, and $C_{LCOE}, C_{LPSP}, and\ C_{REF}$ denote the LCOE, LPSP, and REF values obtained for this evaluation. The first four parameters represent the decisions made by the agent, and the agent's decision directly impacts the last five.

3.5.2.2. Decision-making agent. Starting from the initial state or the updated state, the agent selects the value of the parameters of the previous state. The agent considers the state of the system following his previous decision or the initial state. Consequently, four probability values are predicted between 0 and 1 from the actor of the system (the agent), or a sequence of 4 random values is generated between 0 and 1 based on the value of the epsilon at the time of the decision. The values thus obtained are then recalibrated between the maximum and the minimum to determine the values corresponding to the number of PV panels and WTs, capacity of the battery (BESS), and the capacity of the diesel generator (DG). Here, p denotes the predicted probability value for a parameter, P, and the corresponding value, X, can be defined as follows:

$$X = \left\lfloor X_{\min} + (X_{\max} - X_{\min}) \times \frac{p+1}{2} \right\rfloor \tag{22}$$

where $X_{\min}$ and $X_{\max}$ denote the minimum and maximum values of P.

3.5.2.3. System evaluation. The evaluation of the system involves estimating the amount of energy produced at each moment and the management of the system at each moment over the entire duration (a full year). Based on the state of the system, the LCOE, LPSP, and REF are calculated to determine the impact of each decision. Lastly, the corresponding cumulative agent reward is calculated. This total agent reward, R, is calculated (Eq. (18) without considering P_{grid} in the calculation of the REF and is calculated based on the difference between the demand and production for the HRES (Eq. (15)). The new final state obtained is then added to the buffer for the replay phases.

3.5.2.4. Update states and replay buffer—agent training. When the maximum number of iterations, i.e., the stop condition, is not reached, the system state is updated for the agent with the new values that the agent would have considered as an action and the new cumulative reward based on that decision. The replay buffer, which contains the past decisions and reward history of an agent, is updated with the current decision and the earned reward. The agent is then trained on minibatches of these decisions and the rewards obtained at the end of these past decisions. In particular, a mini-batch of T random transactions is taken from the replay buffer (Eq. (23)).

$$\left\{ \left(s_j, a_j, r_j, s_{j+1}, done_j\right) \right\}_{j=1}^{T} \tag{23}$$

For each of the transactions of T, the target value, y_j, is calculated as shown in Eq. (24), where ζ denotes the discount factor and $Q(s, a;\ \theta^-)$ denotes the target shares value function parameterized by θ^-. θ^- represents the weight of the target network, which is a delayed copy of θ. The model parameters are then updated using the gradient descent method (Eq. (26), where α denotes the learning rate and $L(\theta)$, which is defined by Eq. (25), represents the minimization of the loss function corresponding to θ.

$$y_j = \begin{cases} r_j + \zeta \max\limits_{a'} Q\left(s_{j+1}, a';\ \theta^-\right), done_j = 0 \\ r_j, done_j = 1 \end{cases} \tag{24}$$

$$L(\theta) = \frac{1}{T}\sum_{j=1}^{T}\left(y_j - Q\left(s_j, a_j;\ \theta\right)\right)^2 \tag{25}$$

$$\theta = \theta - \alpha \nabla_\theta L(\theta) \tag{26}$$

Table 6 presents the parameter values used to train the DRL agent.

The replay buffer is used to break the correlations between the sequential data along with the target network, enabling it to stabilize learning (Mnih et al., 2015/02). The agent makes a new decision at the end of this training phase, and the evaluation process is resumed. This process enables the agent to improve their understanding of the system and the environment and conform to it as the training is continued.

3.5.3. Particle swarm optimization

The PSO is a robust and commonly used optimization algorithm. This algorithm was first proposed in 1995 by Kennedy and Eberhart (1995). It is based on the social behavior of birds in flight or fish in schools (Mohamed et al., 2016). Since its inception, PSO has been successfully applied in various fields, addressing a wide range of complex optimization problems. The PSO algorithm was used for the sizing process of the HRES (Kushwaha and Bhattacharjee, 2024; Mansouri Kouhestani et al., 2020). The search process for this problem begins with defining several particles, P, which collaborate during T iterations (generations) to determine the system that satisfies the defined optimality criteria. Each particle has a position $x_i^{(t)}$ at each moment t representing a generation (iteration) (Eq. (28). For the particle, the position constitutes the set of values required to define the system. Essentially, each position in the framework comprises two (PV-BESS) to four (PV-WT-BESS-GEN) values to determine the number of panels, number of turbines, battery capacity, and capacity of the diesel generator based on the system type. The particle also has a velocity $v_i^{(t)}$, which defines the rate of change and the direction when analyzing the search space for each particle (Eq. (27)).

$$v_i^{(t+1)} = \omega v_i^{(t)} + c_1 a_1 \left(\text{pbest}_i - x_i^{(t)}\right) + c_2 a_2 \left(\text{gbest}^{(t)} - x_i^{(t)}\right), \tag{27}$$

$$x_i^{(t+1)} = x_i^{(t)} + v_i^{(t+1)} \tag{28}$$

Generation after generation, each particle stores the values of the parameters required for its solution, $x_i^{(t)}$, in the memory. Thus, the best value is obtained from the goal function and is represented by pbest_i. With each generation, the algorithm updates the best position of all the particles, $\text{gbest}^{(t)}$. The coefficients of inertia, ω, confidence, c_1 and c_2, and random values, a_1 and a_2, enable the algorithm to define the adjustments that must be made to the various parameters, generation after generation. Apart from the decision-making method on the values of the optimization problem parameters, the procedure for evaluating the system remains identical to applying DRL.

PSO is used as the reference method in this study since it is one of the most commonly used methods in HRES optimization (Faria et al., 2023; Bouafia et al., 2023). However, this analysis also includes a discussion of other metaheuristics, such as GA and SSA, along with simulation tools such as HOMER Pro to place the proposed approach in a broader context. The PSO algorithm used here is based on the "global best" algorithm proposed by Miranda (2018).

Table 7 presents the values of some specific PSO parameters assumed in this study.

3.5.4. Multi-objective particle swarm optimization

MOPSO is an extension of the previously reported PSO algorithm. MOPSO was developed specifically for multi-objective, often conflicting, optimization problems, such as the discounted energy cost, REF, and system reliability. MOPSO algorithms depend on the Pareto dominance mechanisms and the use of external archives to store non-dominated solutions. This approach guides the search for well-distributed Pareto fronts, thereby ensuring the diversity of the solutions (Samy et al., 2021; Cui et al., 2022–02). The effectiveness of MOPSO algorithms has been demonstrated in various applications, from standardized test functions to real-life industrial applications, such as industrial process optimization and energy system optimization (Cui et al., 2022–02; Agarwal et al., 2022/05). The results indicate that MOPSO outperforms other evolutionary algorithms in terms of the speed of convergence, quality, and diversity of solutions (Cui et al., 2022–02).

In this study, we applied MOPSO as one of the comparison methods for its ability to generate diversified solution sets that are well

Table 6
DRL train and decision parameters.

Batch size	Learning rate	Policy noise	Policy update frequency	Epsilon decay	Epsilon min	Epsilon
64	0.001	0.2	10	0.995	0.01	0.1

Table 7
PSO Parameter specifications.

Population size	Iteration	Inertia coefficient	Confidence coefficient, c1	Confidence coefficient, c2
10	100	0.9	0.5	0.3

distributed on the Pareto front, as well as its demonstrated performance on benchmarks and real applications (Cui et al., 2022–02; Shu et al., 2023). Fig. 4 depicts the application of MOPSO in this 6-stage research.

In the context of HRES, MOPSO is used as a reference method for simultaneously optimizing the conflicting objectives of the discounted cost of energy (LCOE), energy reliability (LPSP), and renewable energy share (REF). Each particle represents a possible technological configuration of the system (number of PVs, number of WTs, battery capacity, generator capacity) and is developed in the decision space based on its past experiences and those of others via a leader selection mechanism. During the process, MOPSO maintains a dynamic archive of the Pareto optimal solutions, thereby ensuring a good balance between exploration

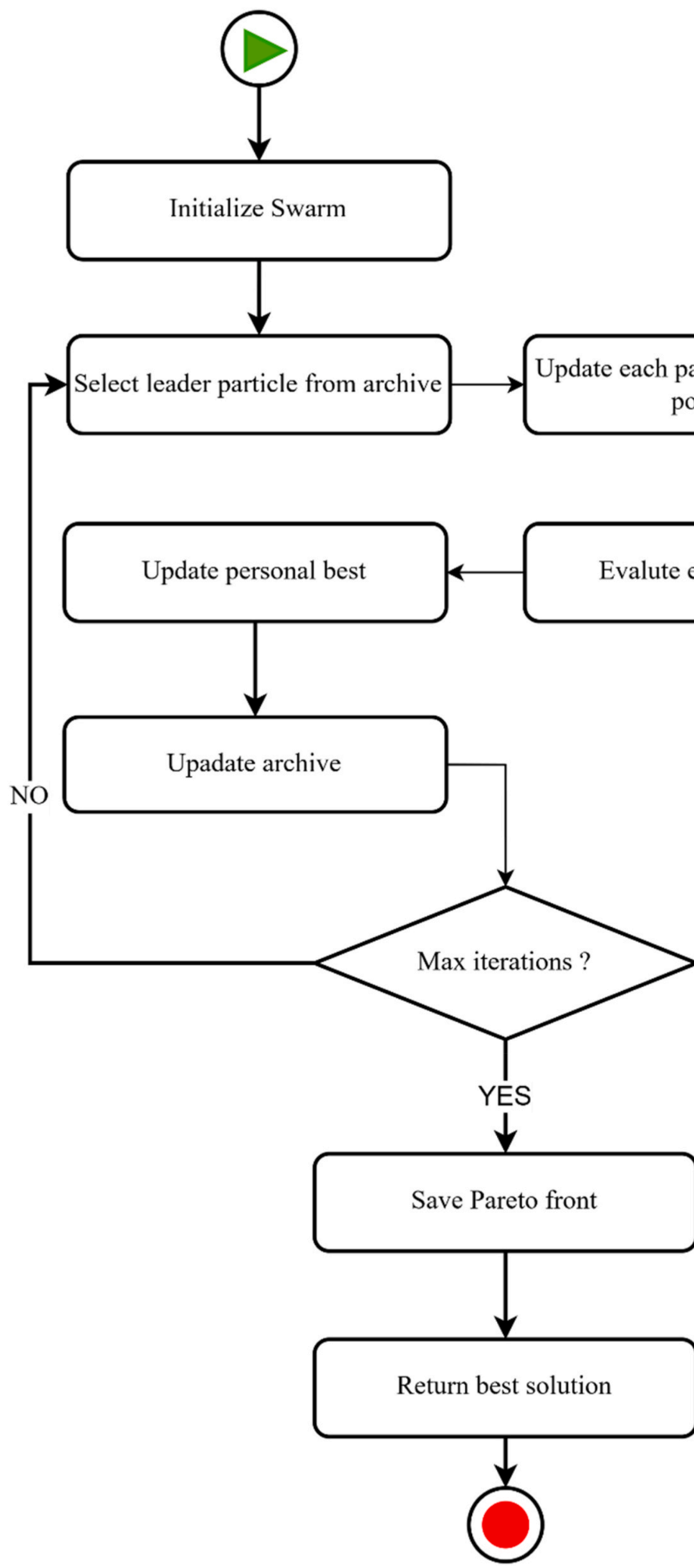

Fig. 4. MOPSO Approach steps.

and exploitation. MOPSO presents a robust benchmark for evaluating the performance of advanced optimization approaches in sustainable energy systems owing to its ability to rapidly converge on various Pareto fronts.

Table 8 presents the values of the MOPSO parameters assumed in this study (Cui et al., 2022–02; Nshimirimana et al., 2021–01).

3.5.5. Non-dominated sorting genetic algorithm II

NSGA-II is a multi-objective evolutionary algorithm that is widely used owing to its ability to efficiently solve multi-objective optimization problems. Proposed by Kalyanmoy Deb in 2002, the excellent performance of NSGA-II was demonstrated in several fields, from engineering and architectural design to water resource management and bioinformatics (Bailey and Caldas, 2023/11; He et al., 2024/07; Shirajuddin et al., 2022–08).

In this study, each individual represents a system configuration (number of PVs, number of WTs, battery capacity, generator capacity), which is evaluated based on the conflicting criteria of this study: LCOE, LPSP, and REF. The NSGA-II algorithm generates a solution through crossover and mutation and classifies the population based on the Pareto fronts and a density distance to preserve the diversity, as shown in Fig. 5. NSGA-II stands out from other multi-objective optimization algorithms owing to its fast, non-dominated sorting, elitism, and diversity without sharing parameters. NSGA-II introduces a more efficient non-dominated sorting method, thereby reducing the computational complexity to O (MN^2), where M denotes the number of objectives and N denotes the population size (Deng et al., 2021–11). Therefore, NSGA-II can identify the reliable trade-offs between the economic, energy, and environmental performance owing to its efficient non-dominated sorting mechanism and its ability to produce a well-distributed Pareto front (Deng et al., 2021–11). Furthermore, its elliptical design combines parental and child populations to select the best configurations, which ensures that high-quality solutions are preserved over generations (Bailey and Caldas, 2023/11; Shirajuddin et al., 2022–08; Bora et al., 2019/01).

Table 9 presents the values of the NSGA-II parameters assumed in this study (Ferreira et al., 2023–05; Wang et al., 2019–05).

4. Results and discussions

This section presents the case studies regarding the energy consumption profiles, type of system, different scenarios, and the imposed constraints.

4.1. Study data

Three energy consumption profiles, P1 [Building 1 consumption], P2 [Building 2 consumption], and P3 [Building 3 consumption], were considered for the significant variability that exists between them from the perspective of total consumption and the average monthly distribution, as shown in Fig. 6 for each of the profiles. These profiles were previously selected at random from the National Renewable Energy Laboratory (NREL) database of energy consumption profiles (Wilson et al., 2021/10). The datasets present hourly consumption profiles over 1 year (8760 rows of values).

Fig. 4 depicts the differences between the different profiles (in kWh). The most important requests occur in December and January for most of these profiles. The periods of low demand are typically recorded in April, May, September, and October.

Table 8
MOPSO Parameter specifications.

Population size	Iteration	Inertia weight	Cognitive coefficient	Social coefficient
20	50	0.5	1.5	1.5

Comparatively, the boxplots in Fig. 7 show that the consumption profile P3 is more stable throughout the year. Additionally, the consumption profiles, P1 and P2, demonstrate significant variability in the summer and peaks in demand in the winter (December and January). The boxplots show that some values in the P1 and P2 profiles could be outliers. A comparison with Fig. 4, which presents the monthly average consumption, indicates that the profile with the highest demand is P1, totaling 72,157.99 kWh, whereas P2 and P3 register 56,341.69 and 18,317.89 kWh, respectively.

This study considers the available Typical Meteorological Year (TMY) data of the region (Wilson et al., 2021/10) since the profiles were randomly extracted from the same region. Table 10 presents a descriptive analysis of this data.

Moderate temperature variability can be observed over the year, with an average value of 14.8 °C and a standard deviation of 10.2 °C. Furthermore, the wind conditions are typically moderate, with an average speed of 3.45 m/s; however, there are a few extreme episodes, with gusts reaching up to 12.8 m/s. The solar irradiance demonstrates a large variation between the sunny and cloudy days, with an average of 169 W/m^2, a maximum value of 993 W/m^2, and a minimum value of 0 W/m^2.

The graph in Fig. 8 demonstrates significant temperature variations in the summer (from June to September), which is a phenomenon of seasonal variations. The minimum temperatures are reached during the winter months (December to February). Meanwhile, the wind speed plot demonstrates significant variations throughout the year, with somewhat larger variations in the winter. The peak periods could represent periods of strong winds, such as storms or gusts. Part 03 of this graph depicts the distribution of solar irradiance over the entire year. A high intensity of solar irradiation can be observed in the summer due to the longer days and direct sunlight, whereas it is minimal in the winter owing to the lower levels of the sun and shorter days. Periods of low intensity could indicate cloudy episodes or storms.

Additionally, the correlation matrix in Fig. 9 depicts the relationships between the different variables (temperature, wind speed, and solar irradiance). The value scale from 0.0 to 1.0 demonstrates the degree of correlation (relationships) between the variables. The darker blue color represents weaker correlation, whereas the darker red color represents stronger correlation, where each variable is strongly correlated with itself and is represented on the diagonal. Sunny days are typically associated with higher temperatures, which presents a positive correlation between the temperature and solar irradiance (0.42). The negative correlation of −0.014 between the wind speed and temperature indicates that windy days tend to be cool. This could explain the windchill factor. Additionally, a positive correlation of 0.27 between wind speed and solar irradiance can be noted. This could be explained by the fact that sunny days, besides having less cloud cover, are typically characterized by moderate wind conditions.

4.2. Study systems

This approach was analyzed using four different types of systems for each profile. Additionally, the analysis was performed with constraints imposed on both the tau (τ) rate of energy permissible by the generator when it is part of the system and the gamma (γ) rate of energy permissible via the electrical grid. The four types of systems are

(a) PV-WT-BESS-GEN
(b) PV-WT-BESS
(c) PV-BESS-GEN
(d) PV-BESS

In each of the above cases, the battery is used for storage when the energy production exceeds the demand at a given time. Consequently, the energy stored in the battery fulfills the energy demand when production cannot meet the demand. Additionally, when it exists, the

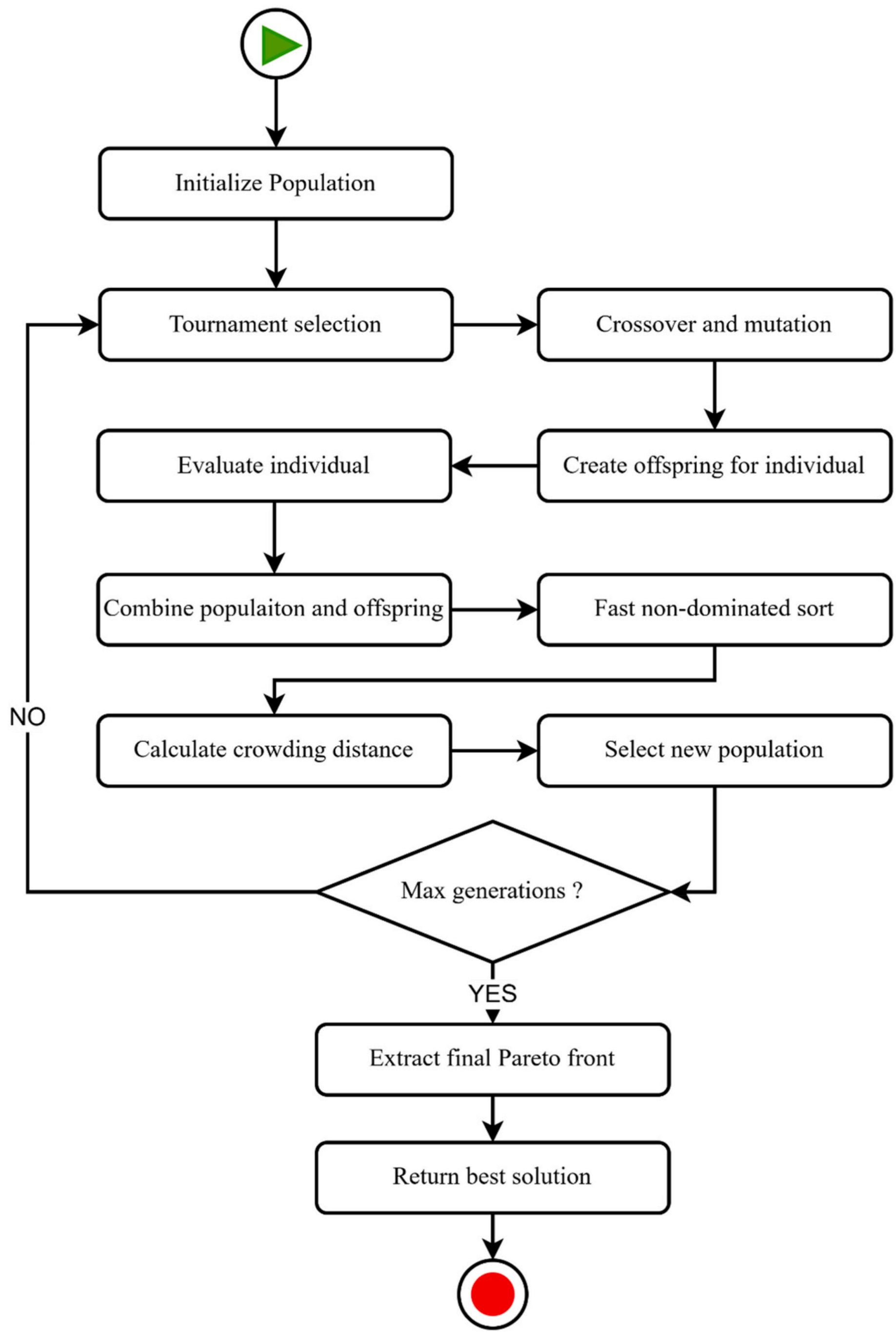

Fig. 5. NSGA-II Approach steps.

Table 9
NSGA-II Parameter specifications.

Population size	Generation	Mutation probability	Crossover probability
20	50	0.1	0.9

generator provides the difference up to the limit imposed when the battery energy cannot sufficiently compensate for the demand.

$\gamma, \tau \epsilon [0.1, 0.2, 0.3, 0.4]$

4.3. Results and discussions

An optimal renewable energy hybrid system is defined based on several factors. A sensitivity analysis demonstrates the behavior of four typical systems on different energy demand profiles under stress to determine the characteristics of the systems that would best meet the energy, cost, and reliability requirements.

Each system was evaluated annually on each profile while varying the energy production constraint from the generator and the constraint of the energy admissible via the grid. This evaluation approach helps in capturing the seasonal and climatic variability. Renewable sources are strongly influenced by seasonality; however, the energy demand of a consumer is also variable between the seasons. This year-long assessment enables these variations to be captured to provide a realistic estimate of the system performance. Table 13 summarizes the best parameter values for each system and profile and the values obtained for each objective.

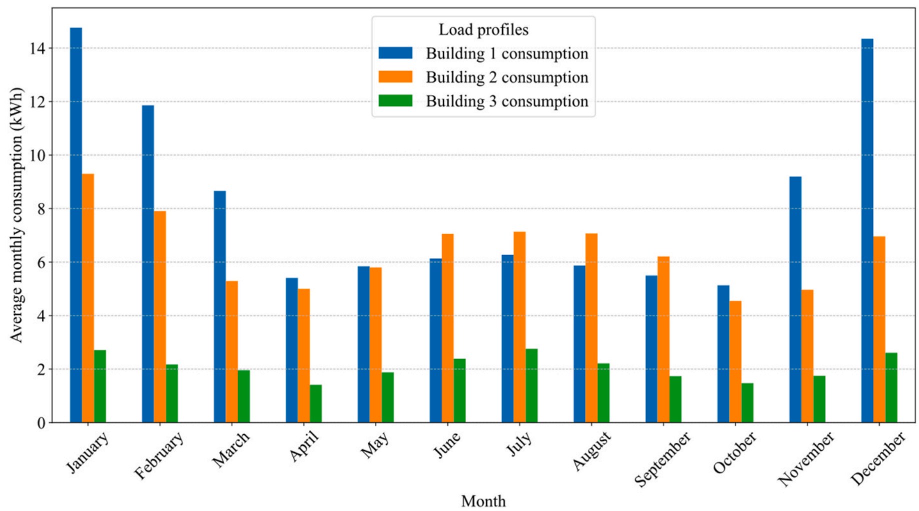

Fig. 6. Average monthly energy consumption.

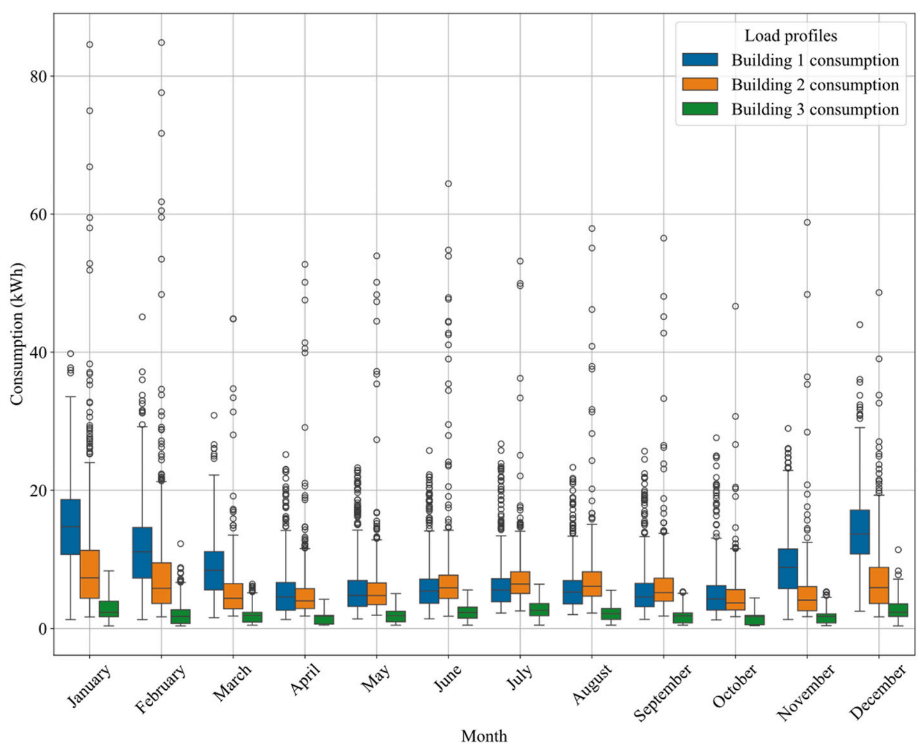

Fig. 7. Profiles comparative boxplots.

Table 10
Descriptive analysis of time series data.

	Temperature [°C]	Wind Speed [m/s]	Solar Irradiance [W/m^2]
Count	8760		
Mean	14.8257	3.4460	169.0066
Std.	10.1880	1.9379	256.6405
Min	−14.0000	0.0000	0.0000
25 %	7.0000	2.1000	0.0000
50 %	16.0000	3.1000	0.0000
75 %	22.8000	4.6000	253.0000
Max	39.0000	12.8000	993.0000

The results presented in Table 13 demonstrate that the P1 profile obtains the lowest energy costs (between 0.2198 and 0.2479 $/kWh) when compared with the P2 and P3 profiles (Fig. 10). This indicates that the specific characteristics of the P1 profile, such as the energy demand and environmental conditions, are better aligned with all the configurations (scenarios studied) of the optimized renewable energy hybrid system. This analysis demonstrates the importance of customizing the energy systems based on the demand profiles and environmental conditions. This approach minimizes the overall costs and the system's cost per kWh. The search intervals typically defined for all the profiles could also explain this result. Based on this analysis, it can be concluded that applying management and optimization strategies to each profile's specificities can considerably reduce the energy expenditure in real-life scenarios.

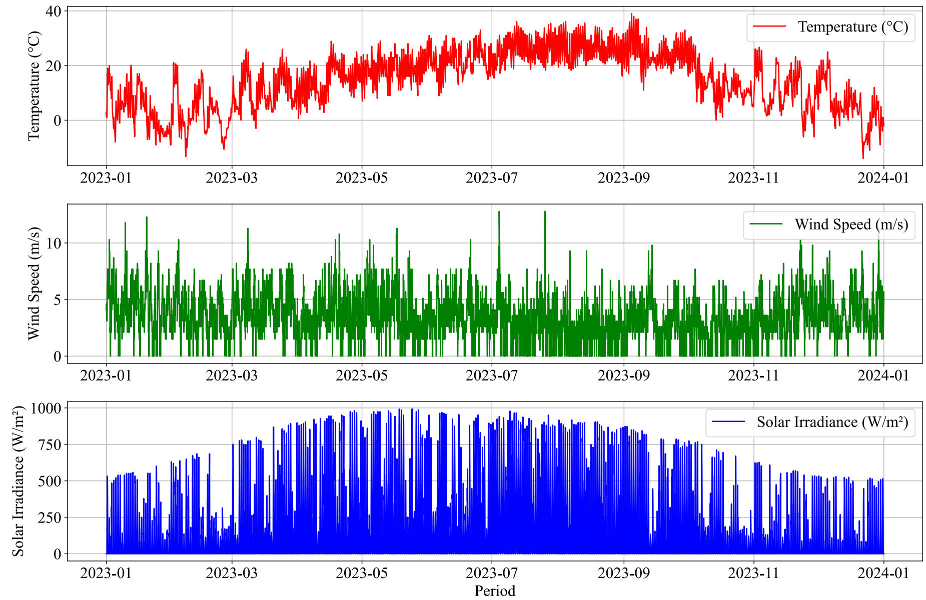

Fig. 8. Time series data.

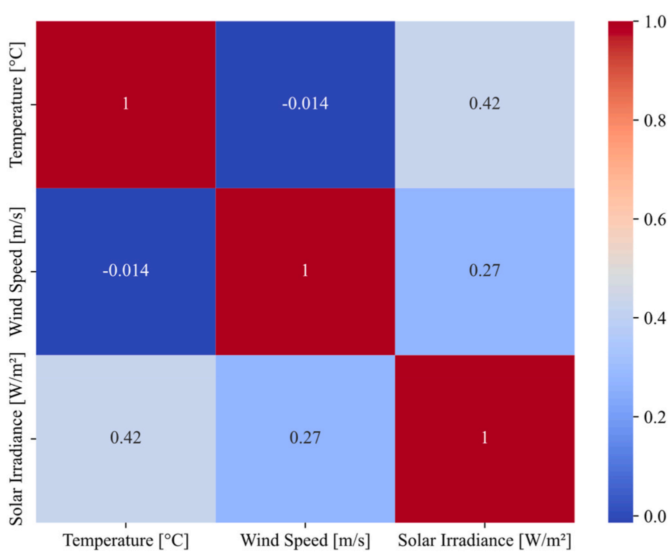

Fig. 9. Correlation matrix of time series data.

The scenario-oriented analysis with and without generators as backup demonstrates that the absence of generators in the system (PV-WT-BESS and PV-BESS) significantly reduces the cost of energy. This indicates that the integration of renewable sources, such as PV panels and WTs, in addition to being beneficial for the environmental footprint, is also economically advantageous in certain cases (Fig. 10). However, this increases the dependence on the intermittent renewable sources, thereby requiring more robust storage sources (BESS) to maintain the system stability (Table 11). This observation concurs with the requirement for an optimized design that comprehensively evaluates the need for a generator based on the variability of demand and available resources.

Additionally, Fig. 10 also shows that scenarios without WTs, such as PV-BESS-GEN and PV-BESS, typically present a higher LCOE. This is more evident for P3, where the costs increase drastically. This analysis indicates an overload on the battery energy storage (BESS) or an oversizing of resources caused by predefined confidence intervals, presenting higher setup and operating costs and less efficient energy management.

This observation corresponds to the definition of discrete values used to determine the optimal values for the generator (between 7.5 and 24 kWh), which may not be adequate for all the energy profiles. The capacity of the generator in each of the scenarios where it exists demonstrates that defining the discrete values for these capacities could significantly affect the total cost of the system and, by extension, the LCOE (the cost per kWh). Defining an inappropriate granularity of values can present additional costs owing to the undersizing or oversizing of the generators based on the actual requirements of a system. Therefore, the values of this complex must be adapted based on the specific characteristics of the consumption profile and the surface area of the building (the available installation space). The proposed approach considers a typical case study, which could present suboptimal results for specific cases. It can also help in minimizing the costs without compromising system performance. This analysis highlights the importance of a carefully calibrated optimization approach.

One significant observation is the improvement in the REF in scenarios that include both the PV panels and WTs (Table 13). This demonstrates the positive impact of these technologies on the sustainability of an energy system by increasing the contribution of renewable sources. This analysis indicates that the PV and WT must be integrated within the same system wherever possible, particularly in contexts where the reduction of CO_2 emissions and dependence on fossil fuels are major issues, to maximize the usage of clean energy.

Fig. 11 shows that the dynamic management of energy flows between the renewable sources, public grid, and generator is crucial. The τ and γ factors could contribute to the definition of adaptive strategies that present a balance between the local exploration and external supply, accounting for the actual demand conditions and the associated costs. The LCOE values (0.63 and 0.58 $/kWh) were obtained in generator-free scenarios for P3. These costs are the lowest energy costs for P3. This indicates that the absence of WT and a high percentage of supply allowed by the generator in that scenario contribute to the increase in the energy cost (0.67 and 0.73 $/kWh). Furthermore, the areas with the lowest LCOE of 0.23 and 0.24 $kWh (blue colors) demonstrate that a moderate proportion of qualifying energy from the generator, combined with a balanced use of qualifying energy via the grid, reduces the energy costs. The energy cost will likely increase when there is a low dependence on the generator, as shown in Fig. 11. This could indicate that when there is insufficient generation, taking energy from the grid is

Table 11
DRL vs. PSO comparison table.

Profile	Method	Scenario	LCOE ($/kWh)
Load Profile 1	DRL	PV-WT-BESS-GEN	0.2282
		PV-WT-BESS	0.2198
		PV-BESS-GEN	0.2479
		PV-BESS	0.2335
	PSO	PV-WT-BESS-GEN	0.2895
		PV-WT-BESS	0.2800
		PV-BESS-GEN	0.2892
		PV-BESS	0.2806
	MOPSO	PV-WT-BESS-GEN	0.3048
		PV-WT-BESS	0.3445
		PV-BESS-GEN	0.4021
		PV-BESS	0.4149
	NSGA-II	PV-WT-BESS-GEN	0.3073
		PV-WT-BESS	0.3712
		PV-BESS-GEN	0.3726
		PV-BESS	0.4457
Load Profile 2	DRL	PV-WT-BESS-GEN	0.2794
		PV-WT-BESS	0.2611
		PV-BESS-GEN	0.3000
		PV-BESS	0.2752
	PSO	PV-WT-BESS-GEN	0.3504
		PV-WT-BESS	0.3323
		PV-BESS-GEN	0.3487
		PV-BESS	0.3319
	MOPSO	PV-WT-BESS-GEN	0.3899
		PV-WT-BESS	0.3786
		PV-BESS-GEN	0.4707
		PV-BESS	0.5465
	NSGA-II	PV-WT-BESS-GEN	0.3836
		PV-WT-BESS	0.4472
		PV-BESS-GEN	0.5004
		PV-BESS	0.5599
Load Profile 3	DRL	PV-WT-BESS-GEN	0.6720
		PV-WT-BESS	0.5826
		PV-BESS-GEN	0.7349
		PV-BESS	0.6264
	PSO	PV-WT-BESS-GEN	0.9490
		PV-WT-BESS	0.8498
		PV-BESS-GEN	0.9269
		PV-BESS	0.8333
	MOPSO	PV-WT-BESS-GEN	1.0364
		PV-WT-BESS	0.9435
		PV-BESS-GEN	1.0539
		PV-BESS	0.8879
	NSGA-II	PV-WT-BESS-GEN	1.0155
		PV-WT-BESS	0.9244
		PV-BESS-GEN	0.9720
		PV-BESS	0.8724

Table 12
Comparison with other research.

Systems/ Resources	Methods	Criteria (Results)	References
PV, WT, BESS	PSO	LPSP (1–30 %)	Mansouri Kouhestani et al. (Mansouri Kouhestani et al., 2020)
PV, WT, BG, DG	MPA, PSO, SSA, GA	LCOE ($0.1799/ kWh)	Kushwaha and Bhattacharjee (Kushwaha and Bhattacharjee, 2024)
PV, WT, BESS, EFCS	Hybrid PSO + GA	waCOE (0.1838 Euro/kWh), DSF (44.63 %), F_{RES} (60.1 %)	Medghalchi and Taylan (Medghalchi and Taylan, 2023/10)
PV, WT, BESS, DG	PSO, epsilon-PSO	LCOE ($0.0148/ kWh), PSRF (0.0730)	Vaka and Matam (Vaka and Matam, 2023)
PV, WT, BESS (LIB/LAB), FC, DG	HOMER Pro	LCOE ($0.295/ $0.284/kWh)	Iqbal et al. (Iqbal et al., 2024/02)

relatively more expensive than obtaining it from the generator. The grid integration improves energy reliability. Additionally, hydroelectricity could be a viable renewable alternative based on the region. Therefore, the proposed approach could be extended to model the impact of different sources of the energy mix of the grid. Furthermore, the difference between the purchase and sale rates of energy significantly affects the overall cost of the system. An increase in the purchase price of electricity tends to increase the dependence on batteries. Conversely, reducing the grid sale price could limit the interest in injecting surplus energy. In future works, sensitivity analysis must be conducted to evaluate these impacts and optimize the energy management.

4.3.1. Comparison between deep reinforcement learning and benchmarks

This section presents a comparative analysis of the results obtained by the proposed DRL method and the three conventional optimization algorithms: PSO, MOPSO, and NSGA-II. The comparison is based on three performance indicators: discounted cost of energy (LCOE), load satisfaction ratio (LPSP), and renewable energy penetration ratio (REF). Simulations were performed on the three different load and generation profiles presented above, representing the energy contexts of increasing complexity.

4.3.1.1. Economic evaluation using LCOE. The results presented in Fig. 12 demonstrate that the DRL method consistently achieves the best LCOE scores for profiles 1 and 2. For profile 1, DRL achieves an average value of 0.23 $/kWh with wide variations, when compared with 0.37 $/kWh for MOPSO and NSGA-II and 0.28 for PSO. This adaptability of DRL can also be observed in profile 2, where DRL maintains a relatively low LCOE of $0.28/kWh, whereas the other methods exceed $0.34/kWh. These results demonstrate the effectiveness of DRL in reducing the costs by learning optimal policies in a dynamic environment.

All the LCOEs increased in profile 3, corresponding to a relatively constant profile. However, DRL maintains a significant advantage (0.65 $/kWh) when compared with 0.98 $/kWh for MOPSO, 0.95 $/kWh for NSGA-II, and 0.89 $/kWh for PSO. This result demonstrates the economic resilience of DRL in unfavorable environments, where the coordinated management of sources is essential.

The results presented in Fig. 13 demonstrate the superiority of DRL in reducing the costs of the system, regardless of the profile and type of the system (PV-BESS, PV-WT-BESS, PV-BESS-GEN, and PV-WT-BESS-GEN).

4.3.1.2. Reliability assessment using LPSP. The DRL method demonstrates a high average LPSP when compared with other methods for profiles 1 and 2 (0.69 and 0.65, respectively), relatively close to the acceptable limit of 0.50, as shown in Fig. 14. This highlights the ability of DRL to provide a system with a reliable power supply. The results obtained by DRL are comparable to those of PSO (0.65 for profile 1 and 0.60 for profile 2), whereas the evolutionary methods (MOPSO and NSGA-II) obtain mean values significantly below the threshold, at approximately 0.40.

In profile 3—where profile incompatibility naturally limits the overall performance of the DRL—PSO, MOPSO, and NSGA-II methods achieve average LPSP values below the threshold. These results demonstrate the ability of the DRL to anticipate load fluctuations and adapt control decisions accordingly.

4.3.1.3. Sustainability analysis using REF. Contrary to the two indicators (LCOE and LPSP), Fig. 15 shows that DRL does not achieve the best performance in terms of the REF for profiles 1 and 2, achieving an average value of 23.25 % and 30.00 %, respectively. Although this seems low, it could be a strategic compromise where DRL prioritizes economic optimization and load satisfaction over the maximum usage of renewables, which may be desirable in some real-life situations where renewable sources are insufficient or unstable.

Table 13
DRL Simulation results.

Profile	Scenario	Parameters				Constraints		Criteria		
		PV (Numbers)	WT (Numbers)	BESS (kWh)	DG (kW)	γ	τ	LCOE ($/kWh)	LPSP (%)	REF (%)
Load profile 1	PV-WT-BESS-GEN	81	23	16.8	7.5	0.4	0.2	0.2282	0.53	33
	PV-WT-BESS	79	25	19.2	–	0.2	–	0.2198	0.69	34
	PV-BESS-GEN	76	–	36	10	0.4	0.1	0.2479	0.68	13
	PV-BESS	76	–	36	–	0.2	–	0.2335	0.87	13
Load profile 2	PV-WT-BESS-GEN	85	24	16.8	7.5	0.4	0.3	0.2794	0.49	42
	PV-WT-BESS	84	24	16.8	–	0.2	–	0.2611	0.65	41
	PV-BESS-GEN	85	–	36	7.5	0.3	0.2	0.3000	0.63	19
	PV-BESS	82	–	31.2	–	0.3	–	0.2752	0.82	18
Load profile 3	PV-WT-BESS-GEN	74	23	19.2	7.5	0.4	0.1	0.6720	0.26	79
	PV-WT-BESS	76	25	16.8	–	0.3	–	0.5824	0.34	80
	PV-BESS-GEN	80	–	33.6	7.5	0.3	0.4	0.7349	0.44	49
	PV-BESS	74	–	36	–	0.2	–	0.6264	0.59	46

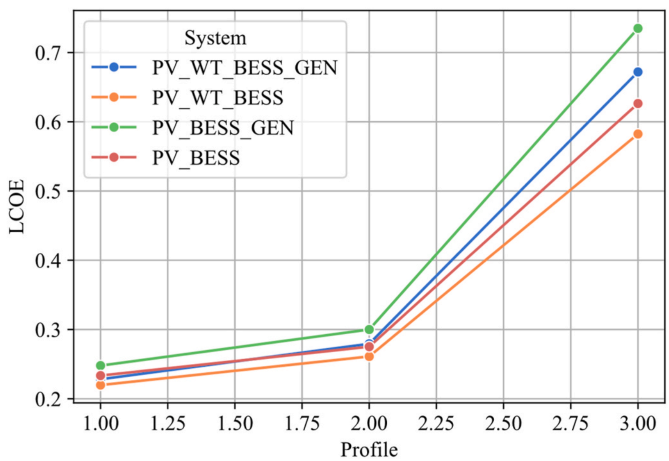

Fig. 10. LCOE vs. Profile for Different Systems.

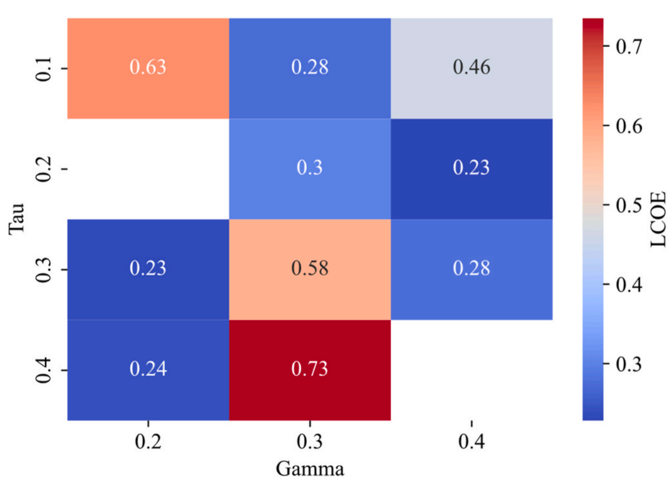

Fig. 11. Heatmap of LCOE

However, in profile 3, DRL achieves an average REF of 63.50 %. This result demonstrates that DRL can effectively integrate renewable sources in contexts where the renewable potential is sufficient and highly exploitable.

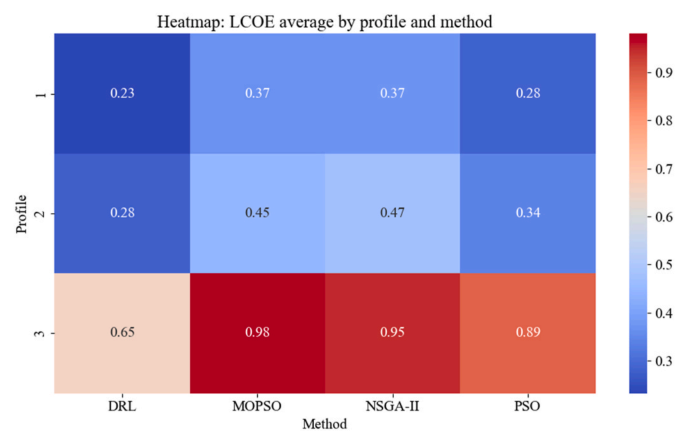

Fig. 12. Heatmap of LCOE average by profile and method.

4.3.1.4. Discussions. The results demonstrate the significant advantages of the DRL method:

- Economic performance: DRL minimizes the LCOE for all profiles and system types, making it particularly suitable in contexts with significant budgetary considerations.
- Reliability: The load satisfaction rates demonstrate that the DRL can maintain an adequate supply, regardless of the context (small, medium, or large load demand).
- Adaptive compromise: Although the REF is sometimes lower, it can be adjusted based on systemic priorities. In this way, DRL can learn multi-objective management policies in real time, unlike conventional evolutionary methods, which are typically limited by a rigid search structure and static heuristics.

Although the PSO, MOPSO, and NSGA-II methods present a high capacity to maximize the use of renewables, they demonstrate a considerable net degradation in terms of the discounted cost of the system. Therefore, the DRL approach, as presented in this study, stands out for its ability to provide a balanced solution between economic performance, operational reliability, and environmental sustainability. This approach is observed to be effective in the dynamic and uncertain environments that are typically observed in HRES owing to its ability to adapt and learn continuously.

The results demonstrate that using DRL for HRES optimization presents a significant reduction in the energy cost, of 21.33 %–30.09 %

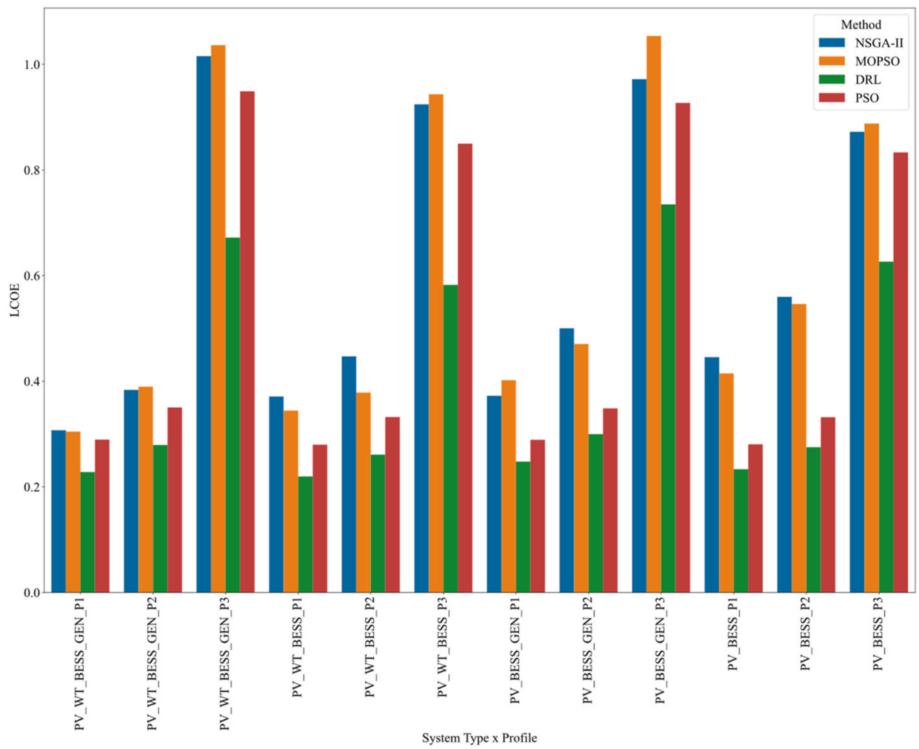

Fig. 13. Comparison of systems by profile and method (DRL vs. PSO vs. MOPSO vs. NSGA-II).

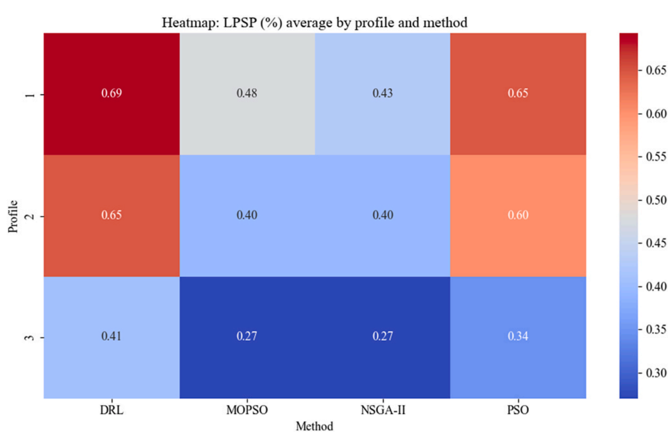

Fig. 14. Heatmap of LPSP average by profile and method.

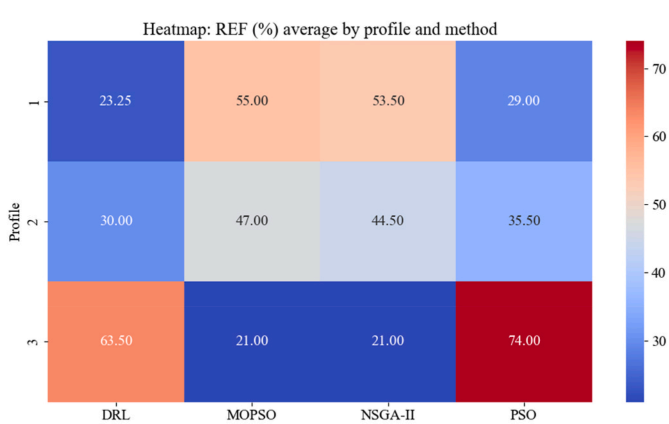

Fig. 15. Heatmap of REF average by profile and method.

when compared with the PSO, of 27.89 %–30.27 % when compared with the MOPSO, and of 27.63 %–28.47 % when compared with the NSGA-II, as shown in Table 7. Dynamic and adaptive learning for optimizing hybrid renewable energy systems could guide future choices in designing and operating more robust, reliable, and economically viable systems.

4.3.2. Comparison of the proposed DRL approach with other works

Several works and various contexts have been considered in this study, such as the maritime transport sector analyzed by Iqbal et al. (Iqbal et al., 2024/02) (Table 12). However, these works evaluate different hybrid system configurations and optimization methods based on several criteria, such as the LCOE, waCOE, REF, and LPSP. Some studies have focused on the PSO and GA methods and their variants.

When compared with the DRL approach, whose energy cost varies between \$0.2198 and \$0.7349/kWh based on the simulated systems and profiles, the proposed approach achieves a minimum LCOE that is better than or almost equivalent to that reported in previous studies (Medghalchi and Taylan, 2023/10; Iqbal et al., 2024/02), which employed the HOMER Pro and PSO + GA approach. However, Vaka and Matam (2023), who performed a 24-h evaluation, reported that optimization over a short period of time could better satisfy the requirements and help in defining a more optimal system.

Medghalchi and Taylan (Medghalchi and Taylan, 2023/10) reported a REF of 60.1 %, which highlights the significant contribution of renewable resources to the energy mix. The proposed DRL approach obtains values between 13 % and 80 % based on the profile and the system, which demonstrates that this approach can define a system that effectively considers the renewable sources.

Additionally, Medghalchi and Taylan (Medghalchi and Taylan, 2023/10) reported that the hybrid PSO + GA method helped in sizing a system that can satisfy almost half of the demand based on the energy generated and stored, with a DSF value of 44.63 %. The DRL approach helps in obtaining a demand satisfaction of approximately 34 % by the locally generated energy in the case of the best LCOE.

Considering the results of Mansouri Kouhestani et al. (Mansouri

Kouhestani et al., 2020), whose approach is based on optimizing the LPSP, the proposed approach obtains high reliability results of between 0.26 and 0.87 %.

Several types of configurations were covered, from simple combinations (PV + WT + BESS) to complex configurations that integrate BG and EFCS. The results of this study demonstrate that the use of advanced technologies such as EFCS or resource technology, such as lithium-ion batteries or lead-acid batteries (Iqbal et al., 2024/02), could improve the cost and reliability indicators (Medghalchi and Taylan, 2023/10; Kushwaha and Bhattacharjee, 2024; Vaka and Matam, 2023). These results present avenues for improving the DRL method by optimizing the operating points of the different technologies integrated into the system.

5. Conclusion

This study demonstrates that DRL can achieve cost-effective, reliable, and highly renewable hybrid energy systems. Four complementary innovations highlight the proposed framework. (1) Adaptive sizing mitigates enumerative searches by letting the agent learn optimal capacities directly through interactions; (2) a cumulative, explicitly multi-objective reward helps in achieving concurrent gains in the LCOE, LPSP, and REF; (3) a simulation-agnostic interface decouples the optimizer from a specific modelling engine, thereby expanding its applicability; and (4) the explicit embedding of generator ramp-rate and grid-quota constraints yields operating strategies that are feasible under stringent real-world limits.

Applied to three highly variable demand profiles, this approach reduces the LCOE by 30.27 %, outperforming PSO, MOPSO, and NSGA-II in all the scenarios. These results demonstrate the potential of DRL as a suitable technique for the dynamic, multi-criteria optimization of HRES.

We presented several graphical illustrations to analyze the results. The experimental results demonstrated the superior adaptability of DRL when compared with the benchmark methods in several scenarios.

Subsequently, the main outcomes of this study can be summarized as follows:

- The DRL reduces the energy costs by 21.33 %–30.27 %, demonstrating its efficiency in minimizing the total costs of the system.
- Renewable energy sources are effectively integrated through the application of DRL, which increases the dependence on sustainable generation over fossil-fuel-based alternatives
- The DRL model dynamically adapts to fluctuations in the energy demand and resource reliability, thereby ensuring greater system resilience.
- The sensitivity analysis demonstrates that variation between the amount of energy obtained from external sources and the renewable energy system affects the overall system profitability, highlighting the importance of economic factors in HRES design.

When implementing optimal HRES, the electricity supplier's permissible injection limit must be considered in the system design.

Although DRL is an effective approach for HRES optimization, it faces certain limitations and presents areas for future improvements. These include:

- Multi-agent reinforcement learning (MARL) for component-level optimization: Rather than using a single DRL agent to manage all the system components, a multi-agent approach could enhance the efficiency by enabling independent agents to optimize the PV panels, WTs, BESS, and backup generators while interacting collaboratively.
- Incorporating tariff policies and grid injection limits: Future works must further analyze regulatory and economic constraints, such as grid energy purchase prices, dynamic tariffs, and restrictions on the renewable energy injection into the power grid. These factors directly affect profitability and system configuration and must be dynamically integrated into the DRL framework.
- Extending the proposed model to alternative configurations: Future works must also analyze additional HRES configurations, including offshore wind power, hydrogen storage, and hybrid systems with thermal energy storage to demonstrate the scalability and robustness of the DRL-based approach across a broader range of applications.

Based on these aspects, the DRL framework can be refined to enhance the intelligence, adaptability, and economic feasibility of HRES, paving the way for next-generation autonomous energy management solutions.

CRediT authorship contribution statement

Inoussa Legrene: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tony Wong:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Louis-A. Dessaint:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

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.

References

18KW all in one volts energies hybrid solar inverter charger | ELIOS. https://volts.ca/products/volts-energies-18k-all-in-one-hybrid-solar-inverter-charger-power-inverters.

Agarwal, P., Agrawal, R.K., Kaur, B., 2022. Multi-objective particle swarm optimization with guided exploration for multimodal problems. Appl. Soft Comput. 120, 108684. https://doi.org/10.1016/j.asoc.2022.108684.

Al-Ghussain, L., Samu, R., Taylan, O., Fahrioglu, M., 2020. Sizing renewable energy systems with energy storage systems in microgrids for maximum cost-efficient utilization of renewable energy resources. Sustain. Cities Soc. 55, 102059. https://doi.org/10.1016/j.scs.2020.102059.

Al-Quraan, A., Al-Mhairat, B., 2024. Sizing and energy management of grid-connected hybrid renewable energy systems based on techno-economic predictive technique. Renew. Energy 228, 120639. https://doi.org/10.1016/j.renene.2024.120639.

Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A., 2017. Deep reinforcement learning: a brief survey. IEEE Signal Process. Mag. 34 (6), 26–38. https://doi.org/10.1109/MSP.2017.2743240.

Bahgaat, N.K., 2023. Estimation of renewable energy systems for mobile network based on real measurements using HOMER software in Egypt. Sci. Rep. 13 (1), 16713. https://doi.org/10.1038/s41598-023-43877-2.

Baidas, M.W., Almusailem, M.F., Kamel, R.M., Alanzi, S.S., 2022. Renewable-energy-powered cellular base-stations in Kuwait's rural areas. Energies 15 (7), 2334. https://doi.org/10.3390/en15072334.

Bailey, E.T., Caldas, L., 2023. Operative generative design using non-dominated sorting genetic algorithm II (NSGA-II). Autom. ConStruct. 155, 105026. https://doi.org/10.1016/j.autcon.2023.105026.

A. Bazzi, H. El Hafdaoui, A. Khallaayoun, K. Mehta, K. Ouazzani, and W. Zörner, "Optimization model of hybrid renewable energy generation for electric bus charging stations," Energies, vol. 17, no. 1, doi: https://doi.org/10.3390/en17010053.

Bora, T.C., Mariani, V.C., Coelho, L.d.S., 2019. Multi-objective optimization of the environmental-economic dispatch with reinforcement learning based on non-dominated sorting genetic algorithm. Appl. Therm. Eng. 146, 688–700. https://doi.org/10.1016/j.applthermaleng.2018.10.020.

Botvinick, M., Wang, J.X., Dabney, W., Miller, K.J., Kurth-Nelson, Z., 2020. Deep reinforcement learning and its neuroscientific implications. Neuron 107 (4), 603–616. https://doi.org/10.1016/j.neuron.2020.06.014.

Bouafia, M., El Fathi, A., El-Hammouchi, A., El Akchioui, N., 2023. Deep analysis on sizing renewable energy system at 12 locations in Morocco using particle swarm optimization. Int. J. Renew. Energy Resour. 13 (3), 1386–1397. https://doi.org/10.20508/ijrer.v13i3.14116.g8811.

Cheraghi, R., Hossein Jahangir, M., 2023. Multi-objective optimization of a hybrid renewable energy system supplying a residential building using NSGA-II and MOPSO algorithms. Energy Convers. Manag. 294, 117515. https://doi.org/10.1016/j.enconman.2023.117515.

Coban, H.H., 2024. A multiscale approach to optimize off-grid hybrid renewable energy systems for sustainable rural electrification: economic evaluation and design. Energy Strategy Rev. 55, 101527. https://doi.org/10.1016/j.esr.2024.101527.

Coban, H.H., Lewicki, W., Brelik, A., 2024. Modeling environmental and economic factors in regional energy optimization. Econ. Environ. 89 (2), 748. https://doi.org/10.34659/eis.2024.89.2.748.

Cui, Y., Meng, X., Qiao, J., 2022. A multi-objective particle swarm optimization algorithm based on two-archive mechanism. Appl. Soft Comput. 119, 108532. https://doi.org/10.1016/j.asoc.2022.108532.

Deng, W., et al., 2021. An enhanced fast non-dominated solution sorting genetic algorithm for multi-objective problems. Inf. Sci. 585, 441–453. https://doi.org/10.1016/j.ins.2021.11.052.

Du, W., Ding, S., 2021. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. Artif. Intell. Rev. 54 (5), 3215–3238. https://doi.org/10.1007/s10462-020-09938-y.

Faria, J., Marques, C., Pombo, J., Mariano, S., Calado, M.d.R., 2023. Optimal sizing of renewable energy communities: a multiple swarms multi-objective particle swarm optimization approach. Energies 16 (21), 7227. https://doi.org/10.3390/en16217227.

Ferreira, B., Antunes, A., Carriço, N., Covas, D., 2023. NSGA-II parameterization for the optimal pressure sensor location in water distribution networks. Urban Water J. 20, 738–750. https://doi.org/10.1080/1573062X.2023.2209553.

He, B.-H., Du, X.-L., Bai, M.-Z., Yang, J.-W., Ma, D., 2024. Inverse analysis of geotechnical parameters using an improved version of non-dominated sorting genetic algorithm II. Comput. Geotech. 171, 106416. https://doi.org/10.1016/j.compgeo.2024.106416.

Iqbal, R., Liu, Y., Zeng, Y., Zhang, Q., Zeeshan, M., 2024. Comparative study based on techno-economics analysis of different shipboard microgrid systems comprising PV/wind/fuel cell/battery/diesel generator with two battery technologies: a step toward green maritime transportation. Renew. Energy 221, 119670. https://doi.org/10.1016/j.renene.2023.119670.

Ismail, M.S., Moghavvemi, M., Mahlia, T., 2013. Techno-economic analysis of an optimized photovoltaic and diesel generator hybrid power system for remote houses in a tropical climate. Energy Convers. Manag. 69, 163–173. https://doi.org/10.1016/j.enconman.2013.02.005.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of ICNN'95-international Conference on Neural Networks. IEEE, pp. 1942–1948. https://doi.org/10.1109/ICNN.1995.488968, 4.

Kiran, B.R., et al., 2022. Deep reinforcement learning for autonomous driving: a survey. IEEE Trans. Intell. Transport. Syst. 23 (6), 4909–4926. https://doi.org/10.1109/TITS.2021.3054625.

Kushwaha, P.K., Bhattacharjee, C., 2024. Socio-techno-economic-environmental sizing of hybrid renewable energy system using metaheuristic optimization approaches. Environ. Prog. Sustain. Energy 43 (3), e14386. https://doi.org/10.1002/ep.14386.

Kushwaha, P.K., Ray, P., Bhattacharjee, C., 2022. Optimal sizing of a hybrid renewable energy system: a socio-techno-economic-environmental perspective. J. Sol. Energy Eng. 145 (3). https://doi.org/10.1115/1.4055196.

Le, N., Rathour, V.S., Yamazaki, K., Luu, K., Savvides, M., 2022. Deep reinforcement learning in computer vision: a comprehensive survey. Artif. Intell. Rev. 55 (4), 2733–2819. https://doi.org/10.1007/s10462-021-10061-9.

Legrene, I., Wong, T., Dessaint, L.A., 2024. Practical cost effectiveness analysis for solar energy systems: case study of stand-alone retail building. In: 2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), pp. 1–5. https://doi.org/10.1109/CPE-POWERENG60842.2024.10604314.

Legrene, I., Wong, T., Mary, N., Dessaint, L.-A., 2025. Reduction in microgrid topology selection time via hybrid branch and bound and k-Nearest neighbors techniques. Mathematics 13 (3), 360. https://doi.org/10.3390/math13030360.

J. Lemaire, R. Castro, and F. Montemor, "Empowering remote and off-grid renewable energy communities: case studies in Congo, Australia, and Canada," Energies, vol. 17, no. 19, doi: https://doi.org/10.3390/en17194848.

Lilienthal, P., 2005. HOMER® Micropower Optimization Model. National Renewable Energy Lab.(NREL), Golden, CO (United States) [Online]. Available: https://www.osti.gov/servlets/purl/860993-NmSLh0/.

Lu, J., Wang, W., Zhang, Y., Cheng, S., 2017. Multi-objective optimal design of stand-alone hybrid energy system using entropy weight method based on HOMER. Energies 10 (10), 1664. https://doi.org/10.3390/en10101664.

Mahmoud, F.S., et al., 2022. Optimal sizing of smart hybrid renewable energy system using different optimization algorithms. Energy Rep. 8, 4935–4956. https://doi.org/10.1016/j.egyr.2022.03.197.

Mansouri Kouhestani, F., et al., 2020. Multi-criteria PSO-based optimal design of grid-connected hybrid renewable energy systems. Int. J. Green Energy 17 (11), 617–631. https://doi.org/10.1080/15435075.2020.1779072.

Medghalchi, Z., Taylan, O., 2023. A novel hybrid optimization framework for sizing renewable energy systems integrated with energy storage systems with solar photovoltaics, wind, battery and electrolyzer-fuel cell. Energy Convers. Manag. 294, 117594. https://doi.org/10.1016/j.enconman.2023.117594.

Memon, S.A., Upadhyay, D.S., Patel, R.N., 2021. Optimal configuration of solar and wind-based hybrid renewable energy system with and without energy storage including environmental and social criteria: a case study. J. Energy Storage 44, 103446. https://doi.org/10.1016/j.est.2021.103446.

Miranda, L.J., 2018. PySwarms: a research toolkit for particle swarm optimization in python. J. Open Source Softw. 3 (21), 433 [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.00433.pdf.

Mnih, V., et al., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533. https://doi.org/10.1038/nature14236.

Mohamed, M.A., Eltamaly, A.M., Alolah, A.I., 2016. PSO-based smart grid application for sizing and optimization of hybrid renewable energy systems. PLoS One 11 (8), e0159702. https://doi.org/10.1371/journal.pone.0159702.

Mokhtara, C., Negrou, B., Settou, N., Settou, B., Samy, M.M., 2021. Design optimization of off-grid hybrid renewable energy systems considering the effects of building energy performance and climate change: case study of Algeria. Energy 219, 119605. https://doi.org/10.1016/j.energy.2020.119605.

Musa, H., Ibrahim, S., 2015. A review of particle swarm optimization (PSO) algorithms for optimal distributed generation placement. Int. J. Energy Power Eng. 4 (4), 232–239 [Online]. Available: https://www.academia.edu/download/83271428/10.11648.j.ijepe.20150404.16.pdf.

Nguyen, T.T., Nguyen, N.D., Nahavandi, S., 2020. Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications. IEEE Trans. Cybern. 50 (9), 3826–3839. https://doi.org/10.1109/TCYB.2020.2977374.

Nshimirimana, R., Abraham, A., Nothnagel, G., 2021. A multi-objective particle swarm for constraint and unconstrained problems. Neural Comput. Appl. 33, 11355–11385. https://doi.org/10.1007/s00521-020-05555-6.

Samy, M.M., Barakat, S., 2019. Hybrid invasive weed optimization - particle swarm optimization algorithm for Biomass/PV Micro-grid power system. In: 2019 21st International Middle East Power Systems Conference (MEPCON), pp. 377–382. https://doi.org/10.1109/MEPCON47431.2019.9008156.

Samy, M.M., Sarhan, H.H., Barakat, S., Al-Ghamdi, S.A., 2018. A hybrid PV-Biomass generation based micro-grid for the irrigation system of a major land reclamation project in Kingdom of Saudi Arabia (KSA) - case study of Albaha area. In: 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), pp. 1–8. https://doi.org/10.1109/EEEIC.2018.8494543.

Samy, M.M., Mosaad, M.I., El-Naggar, M.F., Barakat, S., 2021. Reliability support of undependable grid using green energy systems: economic study. IEEE Access 9, 14528–14539. https://doi.org/10.1109/ACCESS.2020.3048487.

Sandeep, S., Nandihalli, R., 2020. Optimal sizing in hybrid renewable energy system with the aid of opposition based social spider optimization. J. Electr. Eng. Technol. 15, 433–440. https://doi.org/10.1007/s42835-019-00184-z.

Shirajuddin, T.M., Muhammad, N., Abdullah, J., 2022. Optimization problems in water distribution systems using Non-dominated sorting genetic algorithm II: an overview. Ain Shams Eng. J. https://doi.org/10.1016/j.asej.2022.101932.

Shu, X., Liu, Y., Liu, J., Yang, M., Zhang, Q., 2023. Multi-objective particle swarm optimization with dynamic population size. J. Computat. Des. Eng. 10 (1), 446–467. https://doi.org/10.1093/jcde/qwac139.

Tabak, A., Özkaymak, M., Güneser, M.T., Erkol, H.O., 2017. Optimization and evaluation of hybrid PV/WT/BM system in different initial costs and LPSP conditions. Int. J. Adv. Comput. Sci. Appl. 8 (11). https://doi.org/10.14569/IJACSA.2017.081116.

K. N. Ukoima, O. I. Okoro, P. I. Obi, U. B. Akuru, and I. E. Davidson, "Optimal sizing, energy balance, load management and performance analysis of a hybrid renewable energy system," Energies, vol. 17, no. 21, doi: https://doi.org/10.3390/en17215275.

Vaka, S.S.K.R., Matam, S.K., 2023. Optimal sizing of hybrid renewable energy systems for reliability enhancement and cost minimization using multiobjective technique in microgrids. Energy Storag. 5 (4), e419. https://doi.org/10.1002/est2.419.

Wang, Q., Wang, L., Huang, W.-S., Wang, Z., Liu, S., Savić, D., 2019. Parameterization of NSGA-II for the optimal design of water distribution systems. Water. https://doi.org/10.3390/W11050971.

Wilson, E., et al., 2021. End-Use Load Profiles for the U.S. Building Stock. National Renewable Energy Laboratory (NREL). https://doi.org/10.25984/1876417.

Zhu, Z., Lin, K., Jain, A.K., Zhou, J., 2023. Transfer learning in deep reinforcement learning: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 45 (11), 13344–13362. https://doi.org/10.1109/TPAMI.2023.3292075.