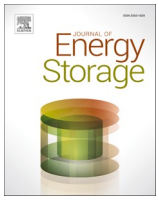
Journal of Energy Storage 131 (2025) 117557

Contents lists available at ScienceDirect

Journal of Energy Storage

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



Research Papers

Solar district energy systems with a seasonal energy storage: Advanced data-driven metaheuristic optimization

Ruslan Kotegov [a], Mohamed Abokersh [b], Carles Mateu [c], Adedamola Shobo [a], Dieter Boer [a], Manel Vallès [a,*]

[a] Universitat Rovira I Virgili, Departament d'Enginyeria Mecànica, Av. Paisos Catalans 26, 43007 Tarragona, Spain
[b] Logitech, Cork, Ireland
[c] GREiA Research Group, University of Lleida, 25001 Lleida, Spain

ARTICLE INFO

Keywords:
Solar district energy system
Thermal energy storage
Life cycle assessment
Multi-objective metaheuristic optimization
Multi-criteria decision making
Sustainability targets

ABSTRACT

Optimizing Solar District Energy Systems (SDES) requires balancing economic feasibility, environmental impact, and computational efficiency. These systems integrate renewable technologies such as solar thermal collectors, photovoltaic (PV) panels, domestic hot water tanks, and seasonal thermal energy storage to meet the heating, electricity, and hot water needs of communities. However, designing cost-effective and sustainable configurations remains challenging due to the system's complexity and competing objectives. To address this, we propose a robust optimization framework that couples TRNSYS simulations with a Python-based control structure, enabling adaptive decision-making and an accurate performance assessment.

Applied to a real SDES case study in Falset, Spain, the methodology identifies system configurations that balance economic and environmental goals. Compared to the fossil-based baseline, the most sustainable solution achieves a 33 % reduction in environmental impact and a 68 % decrease in cost, while the most economical solution lowers environmental impact by 11 % and cuts cost by 88 %. Several scenarios achieve full economic self-sufficiency, with electricity revenues exceeding operating expenses. Although initial investments increase by a factor of 25–32 due to renewable deployment, the optimization ensures strategic allocation to maximize long-term performance and returns.

This hybrid methodology addresses adaptability challenges in energy system design, offering a practical and effective decision-support tool for planners, engineers, and policymakers. It facilitates a comprehensive trade-off analysis between cost and sustainability, unlocking cost-effective pathways for low-carbon urban energy transitions. The proposed methodology improves upon conventional optimization approaches by maintaining simulation accuracy, reducing data requirements, and enhancing adaptability to system changes.

1. Introduction

Integrating solar energy systems into community-based models is important in the evolving landscape of sustainable energy. Energy communities can contribute to expanding renewable energy sources, such as solar, by investing in and owning renewable energy projects. They also can own and operate distribution networks, facilitating the integration of multiple energy generation infrastructures and reducing the need for traditional network upgrades [1]. Solar energy, in turn, has the potential to play a major role in decarbonizing the global energy system by 2050, by providing up to 45 % of the electricity supply and 20 % of the total energy supply [2,3]. This study prioritized solar energy utilization due to its widespread availability and predictability in generation patterns.

Photovoltaic (PV) panels and Solar Thermal Collectors (STCs) reduce reliance on fossil fuels by generating clean energy but still have notable environmental impacts throughout their life cycles. Their production is resource-intensive and often relies on non-renewable energy, indirectly contributing to carbon emissions [4]. In the case of PV energy, while the environmental impact during energy generation is minimal, the impacts associated with the manufacturing process cannot be overlooked [5]. Namely, the inclusion of hazardous materials, such as lead and cadmium in PV panels and other toxic substances in STCs, raises toxicity concerns, especially at the end of their lifecycle, with improper disposal risking environmental and human health [6,7]. Resource depletion from raw

* Corresponding author.
E-mail address: manel.valles@urv.cat (M. Vallès).

https://doi.org/10.1016/j.est.2025.117557
Received 16 September 2024; Received in revised form 30 May 2025; Accepted 27 June 2025
Available online 2 July 2025

Nomenclature

A_{PV}	total aperture area of photovoltaic panels (m^2)
A_{STC}	total aperture area of solar collectors (m^2)
BMF_k	bare module factor of equipment unit k
C	specific water heat capacity (J/kg °C)
$CEPCI$	Chemical Engineering Plant Cost Index
Con_{SST}	purchase cost of the construction material of the seasonal storage tank (€)
D	largest discrepancy between the empirical distributions of the two samples
DVE_k	design variable of equipment unit k
$\dot{m}$	mass flowrate of the recirculating water pumps (kg/s)
α_{CF}	factor of contingency charges and fees
α_k	purchase cost coefficient of equipment unit k
β_k	purchase cost exponent of equipment unit k
δ_d	normalization factor for damage category d
ε_d	weighting factor for damage category d
f_h	fraction concerning the daily total domestic hot water demand (−)
$f_c(x)$	original objective function (RCP or NPC)
$F_f(x)$	cumulative distribution functions of the two samples being compared
FID_d	indicator result for damage category d
IIC	initial investment cost (€)
Ins_{SST}	purchase cost of the insulation material of the seasonal storage tank (€)
NPC	net present cost (€)
OC	operational cost (€)
P	water pump
PCE_k	purchase cost of equipment unit k (€)
Pop	population per municipality (number of people)
PVF_n	present value factor of single future cash flow at the beginning of nth time period (−)
PW_{APP}	electricity cost purchased from the grid (€)
PW_{AUX}	natural gas consumption (€)
PW_{HP}	heat pump electricity consumption (€)
PW_M	maintenance cost (€)
Q_{DHW}	daily domestic hot water demand (J)
RCP	ReCiPe 2016 aggregated impact factor (Pt/MWh)
RC	replacement cost (€)
T_{sup}	supply water temperature (°C)
T_w	tap water temperature (°C)
V	domestic hot water volume per person per day (m^3)
ρ	water density (kg/m^3)
year A	base year
year B	installation year

Indices

c	objective function index
d	damage category
f	dataset index
k	equipment unit

Abbreviations

ADM	Automatic Decision-Making
ANN	Artificial Neural Network
AS	Ant System
BAO	Bat Algorithm Optimization
CDF	Cumulative Distribution Function
CL	Change Logic
COL	Solar collector field
CST	Change Significance Threshold
CV	Cross-Validation
DDDM	Data-Driven Decision-Maker
DHW	domestic hot water
DHWT	domestic hot water tank
EMP	Extended Mathematical Programming
FG	foam glass gravel
FIS	Feature Importance Scoring
FRW	Focused Random Walks
GA	Genetic Algorithm
GASP	Gaussian Sampling Around Point
GBR	Gradient Boosting Regression
GenOpt	Generic Optimization Tool
GWO	Grey Wolf Optimization
HE	heat exchanger
HGAPSO	Hybrid Genetic Algorithm Particle Swarm Optimization
HP	heat pump
HPC	high-performance concrete
HSCSA	Harmony Search-based Chaotic Simulated Annealing
IDG	Initial Data Generation
KS	Kolmogorov-Smirnov goodness-of-fit test
LCA	Life Cycle Assessment
LCC	Life Cycle Costing
LHS	Latin Hypercube Sampling
ML	Machine Learning
MOMVO	Multi-Objective Multi-Verse Optimization
MPSO	Modified Particle Swarm Optimization
MSE	Mean Square Error
MW	mineral wool
NC	normal concrete
P	water pump
PSO	Particle Swarm Optimization
PV	Photovoltaic Panel
PVGIS	Photovoltaic Geographical Information System
RF	Refinement Found
SDES	Solar District Energy System
SEC	specific economic cost
SH	space heating
SM	Simulation Model
SSO	Social Spider Optimization
SST	seasonal storage tank
STC	solar thermal collector
SVM	Support Vector Machine
TOPSIS	Technique for Order of Preference by Similarity to the Ideal Solution
TRNSYS	Transient System Simulation Tool
UHPC	ultra-high-performance concrete
XPS	extruded polystyrene

material extraction and processing further impacts environmental sustainability [8]. Effective recycling and waste management are essential to mitigate these impacts, prevent hazardous material leaching, reclaim valuable materials, and reduce the need for virgin resource extraction. Additionally, installing solar thermal systems can lead to land use conflicts and affect local wildlife. Integrating environmental considerations in deploying these technologies is crucial to minimize their ecological impact [4]. Therefore, though STCs and PVs do not emit greenhouse gases during energy generation, thereby slowing down global warming, reducing dependency on fossil fuels, and enhancing energy security, they still have the potential environmental impacts associated with them [9].

On the other hand, solar energy systems offer a range of financial benefits over conventional energy systems. Coughlin and Kandt [10] highlighted the utility bill savings and reductions in greenhouse gas emissions that come with solar energy systems. Kim et al. [11] conducted a techno-economic analysis of a renewable energy system with solar district heating, showing that it can reduce CO_2 emissions and primary energy savings, and has a lower levelized cost of heat compared to conventional systems. However, NASA [12] noted that the economic benefits of solar energy systems depend on factors such as initial investment, tax treatment, and the cost of conventional energy.

Recognizing the significant financial benefits and potential environmental impact of renewable energy systems, optimizing their deployment and operation to maximize efficiency and minimize environmental footprint is essential. This context creates the basis for a new approach for renewable energy systems optimization.

Thus, even in the realm of renewable energy systems, optimization is indispensable. The inherent variability of energy sources, such as sunlight, combined with ever-changing energy demands and the imperfections in renewable technologies regarding their life cycle necessitate the development of sophisticated optimization strategies. These strategies are crucial for ensuring a reliable, efficient, and sustainable energy supply that can adapt to the dynamic needs of energy communities. Key challenges in this domain encompass managing energy variability, optimizing storage solutions, and maintaining grid stability amid fluctuating supply and demand. To address these challenges, the heuristics approach has been considered as a foundation because "*on the basis of experience or judgment heuristic seems likely to yield a good solution to the problem*" [13] in low computational times but, on the other hand, it is difficult to quantify the quality of the solution. Some limitations of traditional approaches, such as scalability and determinism, point to areas where new techniques can be developed to address these issues. To offset the disadvantages of the heuristic approach, a combination of heuristics and metaheuristics with exact methods in a metaheuristic framework can improve both solution quality and run times [14,15].

To evaluate the benefits of the proposed methodological approach, it is important to consider the most commonly used methods for optimizing renewable energy systems at present. Table 1 outlines the most common methodologies in the field for 2023.

As illustrated, metaheuristic approaches currently dominate the optimization of energy systems, however, a few attempts are still being made to combine heuristics, metaheuristics, and stochastics. This study combines heuristic and metaheuristic methods, as well as stochastic and stratified techniques. In addition to existing ones, the stratified method is powerful yet easy to use and has minimal computational costs as it is not based on machine learning. The proposed combined approach empowers the design and management of renewable energy systems, enabling them to thrive in a dynamic and unpredictable energy landscape.

To complete the picture, it is also necessary to examine the most popular tools for optimizing energy systems. Table 2 presents a concise overview of the fundamental distinctions between the proposed optimization tool in this paper and the most relevant studies in the domain.

In this context, the main advantages over the previous similar studies are distinguished in Table 3.

The approach of the proposed methodology enhances the diversity of the search space and broadens the spectrum of solvable optimization problems by allowing the integration of multiple algorithms in a single execution. This makes the optimization process more efficient and adaptable. The decision not to utilize GetOpt substantially broadens the scope for customizing the optimization procedure. Similarly, bypassing the use of surrogate models can substantially reduce computational expenses, even with major modifications to the original model.

The novelty of the proposed methodology lies in the methodological framework that employs a mix of heuristic rules to rapidly converge toward optimal solutions while avoiding local optima through Particle Swarm Optimization (PSO) [33] (metaheuristic sampling) as well as Latin Hypercube Sampling (LHS) [34] (stratified sampling), Focused Random Walks (FRW) (heuristic sampling), and Gaussian Sampling Around Point (GASP) (stochastic sampling). By bypassing conventional optimization tools, this approach allows for flexible and efficient direct simulations using Python, ensuring high accuracy and adaptability. The methodology successfully balances economic costs and environmental impacts, demonstrating significant advancements over existing optimization strategies in solar district energy systems.

Nevertheless, while the proposed methodology offers notable improvements in flexibility, accuracy, and search efficiency, certain limitations are inherent to the sampling techniques applied. Stratified methods such as LHS provide uniform space coverage but may overlook emergent patterns once optimization progresses [35,36]. Heuristic approaches like FRW risk stagnation near local optima [37–39], while stochastic methods such as GASP can produce redundancy when refinement reaches diminishing returns [40]. Metaheuristic strategies like PSO require careful tuning and offer no formal guarantee of global convergence in complex, high-dimensional landscapes [33,41]. These limitations highlight the necessity of employing a multi-technique approach, where the strengths of each sampling method complement the weaknesses of others, thereby ensuring a more robust and adaptable optimization framework.

Additionally, it is important to acknowledge the general scalability challenge inherent to simulation-based iterative optimization. As highlighted by Amaran et al. [42], the computational burden in simulation optimization increases significantly with the number of decision variables due to the combinatorial growth of the solution space and the absence of gradient information. Similarly, Saltelli et al. [43] emphasize

Table 1
Renewable energy systems optimization methodologies [16].

Methodology	Description	Type
AI-based		
GA	A search that mimics the process of natural selection to find optimal solutions.	Metaheuristic
PSO	An algorithm inspired by the social behavior of birds flocking or fish schooling to optimize problems.	Metaheuristic
MPSO	An enhanced version of PSO with modifications to improve convergence and solution accuracy.	Metaheuristic
GWO	A nature-inspired algorithm based on the social hierarchy and hunting behavior of grey wolves.	Metaheuristic
BAO	An algorithm that uses the echolocation behavior of bats to find optimal solutions.	Metaheuristic
AS	A probabilistic technique inspired by the behavior of ants seeking paths to food is useful for solving optimization problems.	Metaheuristic
EMP/ANN	Combines advanced mathematical programming with artificial neural networks for enhanced problem-solving capabilities.	Deterministic, surrogate model
HYBRID		
HGAPSO	Combines genetic algorithms and particle swarm optimization for enhanced search capabilities.	Metaheuristic
GA, Nelder-Mead simplex, Improved HSCSA	Integrates GA, Nelder-Mead, and HSCSA for robust and efficient optimization.	Heuristic, metaheuristic, stochastic
SSO	A technique inspired by the cooperative behavior of social spiders in nature.	Metaheuristic
MOMVO	Solves multi-objective problems by simulating the interaction between multiple universes.	Metaheuristic

Table 2
Comparison of the proposed methodology with the most relevant studies.

Ref	Main tool	Optimization tool	Data processing tool	Extensive data generation	Surrogate model	N⁰ of samples	Publication date
[17]	TRNSYS	manual runs	n.m.*	No	No	–	2012
[18]	TRNSYS	GenOpt	TRNSYS	No	No	–	2016
[19]	TRNSYS	GenOpt	n.m.	No	No	–	2016
[20]	TRNSYS	SVM	MATLAB	Yes	Yes	up to 4096	2018
[21]	TRNSYS	GenOpt	TRNSYS	No	No	–	2019
[22]	TRNSYS	GenOpt	n.m.	No	No	–	2019
[23]	TRNSYS	ANN	MATLAB	Yes	Yes	2048	2020
[24]	TRNSYS	GenOpt	n.m.	No	No	–	2020
[25]	TRNSYS	GenOpt	n.m.	N.m.	N.m.	–	2021
[26]	TRNSYS	GenOpt	n.m.	No	No	–	2021
[27]	TRNSYS	GenOpt	n.m.	No	No	–	2023
[28]	TRNSYS	n.m.	n.m.	No	No	–	2023
[29]	TRNSYS	MATLAB	MATLAB	Yes	No	80,000	2024
t.s.**	TRNSYS	Python	Python	No	No	–	–

* n.m. – not mentioned.
** t.s. – this study.

Table 3
Comparative advantages of TRNSYS-Python integration over TRNSYS-GenOpt and TRNSYS-Surrogate models.

Feature	Advantages of TRNSYS–Python framework	GenOpt limitations [30]
Enhanced flexibility	Extensive libraries for custom development tailored to specific needs.	Allows only a predefined set of algorithms and lacks runtime flexibility for changing optimization strategies.
Advanced data-handling	Superior data manipulation and processing for large datasets.	Does not offer internal data processing or post-processing modules; users must handle data externally.
Machine learning integration	Supports advanced algorithms for predictive modeling and optimization.	Has no built-in support for machine learning integration or adaptive decision-making techniques.
Reduced computational demand	Heuristic and adaptive sampling techniques minimize redundant evaluations and concentrate search efforts on promising regions, improving convergence speed.	Although GenOpt's individual algorithms are adaptive, the platform does not support runtime integration or switching between multiple techniques, which may lead to suboptimal convergence paths or require trial-and-error to identify the most suitable strategy.
Versatile optimization strategies	Integrates various techniques to avoid local optima and explore the solution space thoroughly.	Supports one optimization technique per run and cannot dynamically switch between strategies during execution.
Customizability and source code access	Full control over scripting and customization with open-source access.	Limited to XML-based configuration, offering minimal customization capabilities beyond its core functionalities.
Enhanced logging and error handling	Comprehensive logging and error handling for real-time tracking and debugging.	Lacks detailed logging and diagnostic tools; users must create external workarounds to monitor performance and detect anomalies.

Feature	Advantages of TRNSYS–Python framework	Surrogate model approach limitations [31,32]
Direct simulation accuracy	Uses direct TRNSYS simulations for the highest accuracy and fidelity.	Approximate system behavior and may misrepresent results outside the trained dataset.
No need for extensive data generation	Effective with small datasets, reducing preparatory workload and speeding up optimization.	Require extensive pre-training datasets before optimization can begin.
Adaptability to model changes	Adapts seamlessly to model updates without retraining, incorporating changes dynamically.	Must be retrained when the system structure or parameters change significantly.

that the computational cost of model exploration rises sharply with dimensionality, as each additional degree of freedom amplifies the number of required evaluations to ensure adequate coverage of the design space. Although the present study successfully demonstrates optimization with four design variables, extending the framework to higher-dimensional systems may require additional methodological enhancements – such as surrogate-assisted refinement, dimensionality reduction, or parallel computing – to maintain tractability and solution quality.

Beyond the methodological considerations, it is equally important to contextualize the practical implementation of the proposed framework. The Solar District Energy System (SDES) analyzed in this study serves as a representative application of the optimization methodology. Designed to provide space heating, domestic hot water, and electricity to residential communities, the system integrates renewable energy technologies alongside a Seasonal Storage Tank (SST) to enhance efficiency and reliability. With advanced control strategies enabling scalable and demand-responsive operation, the SDES supports the transition toward sustainable energy communities. Its deployment offers not only environmental benefits but also socioeconomic advantages such as reduced energy bills, enhanced energy autonomy, and job creation [9,44].

The main objective of this study was to optimize the design and functioning of the SDES to achieve both economic and environmental efficiency using the proposed optimization framework applied to a real case study.

The remainder of this manuscript is organized as follows: Section 2 presents the proposed methodological framework, detailing the optimization architecture and the integrated TRNSYS-Python interface. Section 3 introduces the case study, including climate conditions, system specifications, and economic/environmental data. Section 4 discusses the results, including model performance, optimization outcomes, and economic and environmental analysis. Finally, Section 5 provides the key conclusions and perspectives for future research.

2. Methodology

For the attainment of objectives, we undertook multi-objective optimization for the SDES through predictive modeling, focusing on reducing the system's net present cost and its environmental footprint. Concurrently, we monitored the system's performance to ensure optimal

performance for a residential community.

At the core of the proposed methodology lies energy modeling, a fundamental concept in the domain of renewable energy systems that is essential for understanding and optimizing the energy processes therein. This was facilitated by the use of the Transient System Simulation Tool (TRNSYS) [45] integrated with Python. The approach leveraged the "subprocess" Python library [46] to pass all the necessary variables to TRNSYS and perform the simulation. This combination afforded exploration of diverse system configurations, transcending geographical constraints and accommodating varied system characteristics while enabling effective solutions for renewable energy systems.

2.1. Data generation for optimization

In this research, we assessed the economic aspects of the system setup using the Life Cycle Costing (LCC) approach following the Tulus et al. [18] study that was carried out based on Duffie and Beckman (2013) [47] and Kalogirou (2009) [48] methodology. Alongside evaluating the system's economic viability through LCC, its environmental footprint is determined through the Life Cycle Assessment (LCA) methodology [49].

2.1.1. Data preparation for economic analysis

The LCC methodology provides a comprehensive framework for assessing the economic efficiency of an energy system. It takes into account the initial investment, operating, and replacement costs throughout the system's life cycle. This approach allows decision-makers to gain a holistic understanding of the system's financial performance over time. Considering all these factors helps to minimize unnecessary expenditures during the design phase, even though there may be an upfront investment required [50].

At the heart of the LCC method lies the concept of future system value, which involves discounting all costs incurred over the system's lifetime back to their present value. Net Present Cost is calculated following standard engineering economics practices [51], incorporating initial investment (IIC), recurring operational expenses (OC), and future replacement costs (RC) discounted over the system's lifetime:

$$NPC = IIC + OC + RC \tag{1}$$

The IIC encompasses the upfront capital expenditure at the onset of the project, covering costs for equipment purchase, installation, and transportation. This also includes provisions for any unforeseen expenses, structured as follows:

$$IIC = (1 + \alpha_{CF}) \sum_{k} (\mathrm{PCE}_k \bullet \mathrm{BMF}_k) \tag{2}$$

where PCE_k represents the initial capital cost of purchased unit k, BMF_k is the bare module factor, which states the installation and transportation cost of unit k, and α_{CF} donates for the contingency factor.

The capital cost of purchased equipment (PCE_k) is updated using the Chemical Engineering Plant Cost Index (CEPCI) (Eq. (3)); PCE_k fetched the initial purchase cost from the base year (year A) to the year of installation (year B), it is a widely accepted method in process and energy system cost estimation [52].

$$\mathrm{PCE}_k = \mathrm{PCE}_k^{\text{year A}} \bullet \frac{\mathrm{CEPCI}^{\text{year B}}}{\mathrm{CEPCI}^{\text{year A}}} \tag{3}$$

The initial purchase cost of the unit k at year A can be estimated with the following equations:

$$\mathrm{PCE}_k^{\text{year A}} = \alpha_k \bullet \mathrm{DVE}_k^{\beta_k} \quad \forall k = \mathrm{COL}, \mathrm{DHW}, \mathrm{AUX} \tag{4}$$

$$\mathrm{PCE}_k^{\text{year A}} = \mathrm{DVE}_k^{\beta_k} \bullet 10^{\left[\alpha_k \left(log_{10}(\mathrm{DVE}_k)\right)^{\beta_k}\right]} \quad \forall k = \mathrm{HE}_1, \mathrm{HE}_2, \mathrm{HE}_3 \tag{5}$$

$$\mathrm{PCE}_k^{\text{year A}} = \alpha_k \bullet ln\left(\frac{\mathrm{DVE}_k}{1000}\right) + \beta_k \quad \forall k = \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3, \mathrm{P}_4 \tag{6}$$

$$\mathrm{PCE}_k^{\text{year A}} = Ins_{\mathrm{SST}} + Con_{\mathrm{SST}} \quad \forall k = \mathrm{SST} \tag{7}$$

where:

$$Ins_{\mathrm{SST}} = \alpha_k \bullet \mathrm{DVE}_k^{\beta_k} \quad \forall k = \mathrm{XPS}, \mathrm{MW}, \mathrm{FG} \tag{8}$$

$$Con_{\mathrm{SST}} = \alpha_k \bullet \mathrm{DVE}_k^{\beta_k} \quad \forall k = \mathrm{NC}, \mathrm{HPC} \tag{9}$$

$$Con_{\mathrm{SST}} = \alpha_k \bullet e^{\left(\frac{\beta_k}{10^5}\mathrm{DVE}_k\right)} \quad \forall k = \mathrm{UHPC} \tag{10}$$

Here, α_k and β_k represent the parameters for purchasing cost of equipment or materials, and CAP_k denotes the design variables for equipment k. The optimization process considered independent design variables such as STC area (A_{STC}), SST and domestic hot water tank (DHWT) volumes (V_{SST}, V_{DHWT}), and the PV panels area (A_{PV}). Their optimization ranges are determined by generating a feasible set of scenarios using the TRNSYS model, these scenarios are used to obtain the minimum and maximum size of each system component. Variables such as heat exchangers (HE_1, HE_2, HE_3), and the five pumps with their respective mass flow rates ($\dot{m}_1, \dot{m}_2, \dot{m}_3, \dot{m}_4, \dot{m}_5$) along with the Heat Pump (HP) and the Auxiliary Heaters (AUX_1, AUX_2), are categorized as dependent, with their specifications fine-tuned in line with electricity, space heating, and domestic hot water demands, all dependent on the primary design variables through specific mathematical formulations. The remaining variables, including the insulation types for SST (XPS, MW, FG, indicating extruded polystyrene, mineral wool, and foam glass gravel) and its construction materials (NC, HPC, UHPC, denoting normal concrete, high-performance concrete, and ultra-high-performance concrete), are predetermined as per the study by Abokersh et al. [53].

The OC includes yearly maintenance costs for different types of equipment, along with expenses related to electricity consumption and the use of auxiliary natural gas (AUX) heaters, structured as follows:

$$OC = \mathrm{PW_M} + \mathrm{PW_{APP}} + \mathrm{PW_{AUX}} + \mathrm{PW_{HP}} \tag{11}$$

where $\mathrm{PW_M}$, $\mathrm{PW_{APP}}$, $\mathrm{PW_{AUX}}$, $\mathrm{PW_{HP}}$ stand for the annual maintenance cost, electricity purchased from the grid, natural gas consumption, and HP electricity consumption cost respectively.

The RC accounts for the depreciation of key equipment over the lifespan of the SDES, including the STC, DHWT, heat exchangers (HEs), AUXs, and water pumps, calculated as follows:

$$RC = \mathrm{PVF}_n \sum_{k} (\mathrm{PCE}_k \bullet \mathrm{BMF}_k) \tag{12}$$

where PVF_n is the present value factor of future cash flow at year n.

The economic parameters of SDES equipment can be found in Section 3 "Case study".

2.1.2. Data preparation for environmental analysis

The LCA methodology provides a comprehensive evaluation of the entire lifecycle of a product or system, from its production to its disposal. This approach follows the "cradle-to-grave" principle [54], which means that every stage of the product's lifecycle is taken into account. According to the ISO 14040 standards [55–57], the LCA process consists of several phases. First, the goal and scope of the assessment are defined. Then, a life cycle inventory is compiled, which includes all the inputs and outputs of the system throughout its life. Finally, an impact assessment is conducted to evaluate the environmental and social impacts of the product.

For this study, the initial phase of LCA establishes the system boundary and functional unit, adopting a "cradle to gate" perspective that omits the end-of-life disposal phase. This approach is typical in LCA

studies, where the disposal phase's impact is often considered minimal relative to the production and usage phases. Similarly, the recycling phase is not emphasized due to its complexity and the challenges associated with accurate forecasting [58].

During the life cycle inventory and impact assessment phase, data on material inputs and outputs, along with associated energy consumption, are collected from various databases for the duration of the system's construction and operation. This analysis included the manufacturing impacts of equipment and the SDES's consumption of utilities such as natural gas and electricity over its entire lifetime. Additionally, the transportation of materials to the construction site and the operational impact of system components are evaluated, with impact data sourced from the Ecoinvent 3.9.1 database [59].

Subsequently, the compiled inventory data are categorized and assessed for their impact on human health, ecological systems, and resource depletion. In this study, the ReCiPe 2016 framework is utilized to assess the environmental impacts of integrating the main components: STC, HP, SST, and PV into the SDES as well as HEs and water pumps. Rather than relying on mid-point indicators, an aggregated endpoint indicator metric (RCP) is employed for a more accurate reflection of the SDES's environmental performance across varying system sizes. This metric allows for a clearer interpretation of the overall environmental effects. The *RCP* can be expressed as follows:

$$RCP = \sum_{d} \delta_d \bullet \varepsilon_d \bullet \mathrm{FID}_d \quad \forall d \tag{13}$$

where FID_d denotes the final impact of the damage category d, and δ_d, ε_d represent the specific normalization and weighting factors. The normalization factors are determined based on the impacts on European land uses and associated material extractions [60]. Conversely, the weighting factors are derived from guidelines provided in the ReCiPe 2016 framework [21].

The environmental impact of SDES components and materials can also be found in the "Case study" section.

2.2. Optimization framework

The main goal of the optimization procedure is to reduce the total

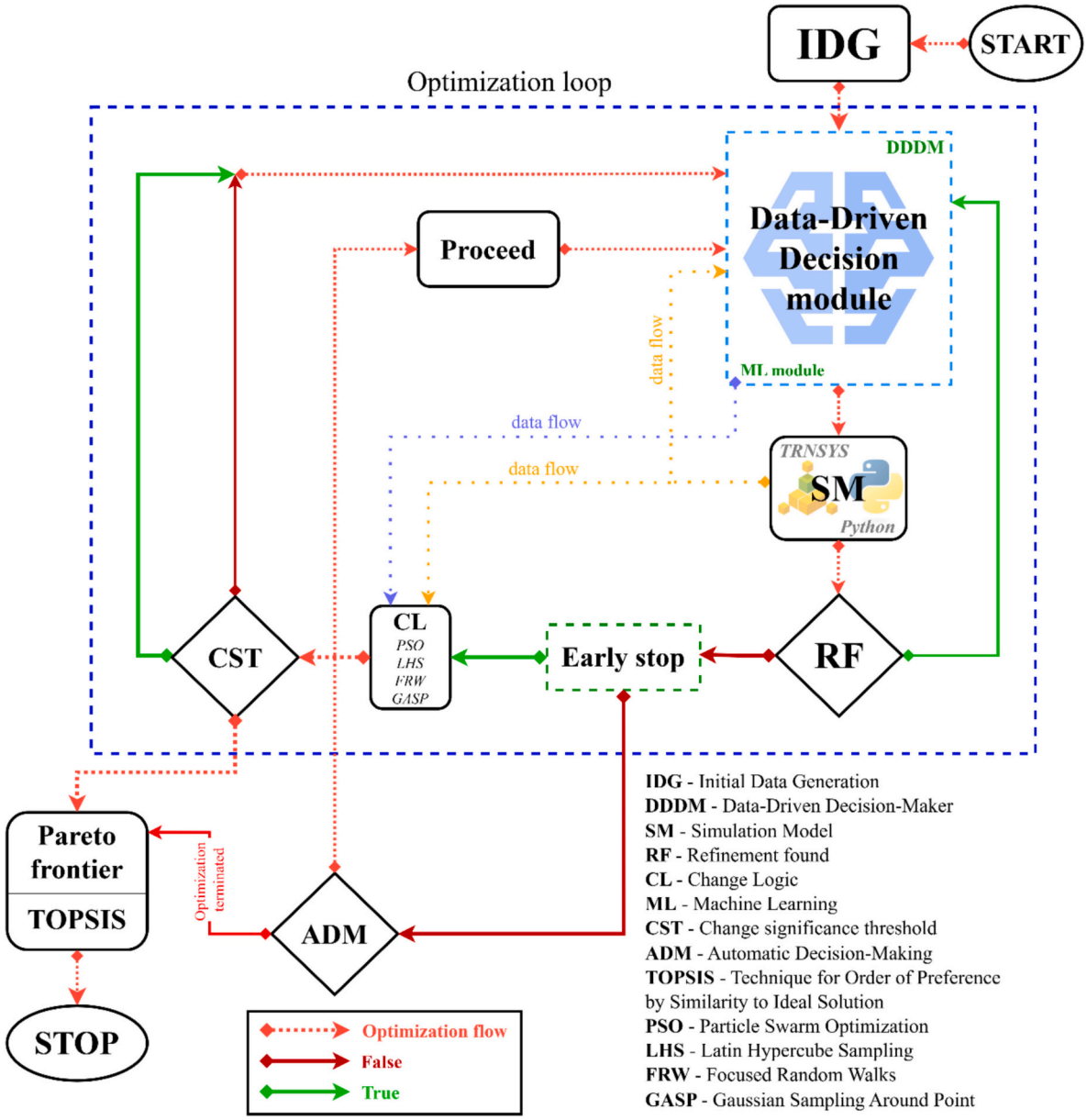


Fig. 1. Optimization framework.

SDES cost (NPC) and its environmental impact (RCP) (objective functions) simultaneously while still satisfying the technical requirements. The proposed general algorithm for the sustainable assessment of a system is outlined in Fig. 1. The full-feature optimization framework is provided in the annex, Fig. 1.

The current section delineates the sequential steps of an efficient optimization framework designed to ascertain the optimal configuration for sustainable energy systems. The methodology leverages Data-Driven Decision-Making (DDDM) to hone in on the most influential variables using Feature Importance Scoring (FIS) and k-fold cross-validation (CV) within a machine-learning context. For the proposed SDES system, the objective function is to optimize the NPC, and RCP is given as:

$$\begin{aligned} &min && \{f_1(x), f_2(x)\} && \\ &s.t. && h_i(x) = 0, && \forall i \in \{1,\ldots,m\} \\ & && g_j(x) \geq 0, && \forall j \in \{1,\ldots,p\} \\ & lb_q \leq x_q \leq ub_q && \forall q \in \{1,\ldots,n\} && \end{aligned} \tag{14}$$

where x is the vector of design variables, f_1 is RCP and f_2 is NPC $h_i(x)$; $h_i(x)$ and $g_j(x)$ are the equality and inequality constraints, respectively, that must be simultaneously satisfied. The equality constraints correspond to the physics of physical systems solved in TRNSYS (e.g., energy balances), and inequality constraints correspond to the technical evaluation of SDES. lb_q and ub_q are the lower and upper bounds, respectively, for each variable x_q; m, p, and n are the total number of equality and inequality constraints, and the total number of design variables constraints, respectively. Through these constraints, annual solar efficiency is maintained $> 60\%$, and renewable energy fraction for thermal energy is maintained $> 50\%$ as recommended by Bauer et al. [61] and Tecnalia [62].

2.2.1. Initial Data Generation (IDG)

The optimization process commences with the generation of an initial data set derived from the extended predefined dataset of system designs. This generated set comprises a series of multi-dimensional points (system designs) where each point encapsulates a distinct combination of independent and dependent variables. The independent variables include the STC and PV areas, as well as SST and DHWT volumes. The dependent variables (objective functions) are RCP measured in ReCiPe 2016 points per MWh and NPC in M€. LHS was chosen for IDG since it ensures uniform extraction of designs, providing a well-distributed and effective starting point for optimization by identifying potential areas of interest.

To ensure the adequacy of the initial LHS for identifying potential areas of interest, we compared the distribution of the LHS samples with the full dataset using the Kolmogorov-Smirnov (KS) goodness-of-fit test [63]. The LHS was generated for 50 samples from a database of approximately 10,000 system designs, ensuring uniform coverage of the design space. The following equation defines the KS test statistic:

$$D = \sup_x |F_1(x) - F_2(x)| \tag{15}$$

where $F_1(x)$ and $F_2(x)$ are the cumulative distribution functions (CDFs) of the two samples being compared. The empirical distribution function is a step function that increases by $1/n$ at each of the n sample points; and $\sup_x$ is a supremum over x means we are looking for the maximum value of the absolute difference between the two empirical distribution functions across all values of x.

In the context of the KS test, Eq. (15) was applied for each variable (STC area, SST volume, DHWT volume, PV area) to compare the distribution of the LHS samples with the full dataset. The KS statistic D thus represents the largest discrepancy between the empirical distributions of the two samples, providing a measure of how well the LHS sample represents the full dataset, the p-value for each variable was computed as part of the KS test.

The KS test outputs are presented in Table 4.

Table 4
Results of the KS test applied for each variable.

Variable	KS statistic	P-value
STC area (m^2)	0.100000	0.665832
SST volume (m^3)	0.090909	0.772087
DHWT volume (m^3)	0.200000	0.032175
PV area (m^2)	0.058824	0.991443

These results indicate that for the STC area, SST volume, and PV area, the LHS samples are uniformly distributed and statistically similar to the full dataset, as evidenced by the low KS statistics and high p-values. However, the DHWT volume variable shows a significant difference, with a higher KS statistic and a p-value below the significance level of 0.05, suggesting that additional refinement may be necessary for this parameter. Overall, the KS test shows that the 50 samples are the effective initial dataset for the optimization process, particularly for most variables under consideration.

2.2.2. Data-Driven Decision-Making (DDDM)

As an essential sub-module within the DDDM, the FIS analyzes the input data to determine the significance of each independent variable to the dependent outcomes.

FIS is a machine learning technique that helps identify and rank the importance of different input variables in a model's predictions. It is used in models like Gradient Boosting Regression (GBR) to determine how each feature contributes to the accuracy of the predictions. By assigning scores to each feature, this method allows us to prioritize which features have the most significant impact on the dependent variable. This approach makes it easier to optimize models more efficiently by focusing on the most influential features. It also helps simplify complex models by identifying less significant features that can be altered or ignored without significantly affecting predictive performance.

In short, using regression analysis, the FIS assigns a numerical value to each variable, creating a ranking of importance. A higher FIS score means that the variable has a stronger impact on the objective functions (environmental and economic indicators of the system).

The DDDM module synthesizes the FIS outcomes with CV practices to mitigate overfitting risks. By concentrating on the most impactful variables identified by the FIS, the DDDM module guides the optimization trajectory, ensuring an effective and expedient approach to solution refinement.

2.2.3. Simulation Model (SM)

The SM module acts as the core execution interface between the optimization routine and the TRNSYS environment. It receives input files generated by Python, runs the energy performance simulations in TRNSYS, and returns structured outputs for evaluation. Additionally, the SM handles preprocessing tasks, such as grouping input files by simulation runs, and postprocessing tasks, such as aggregating simulation results into unified datasets for refinement checks and decision-making modules. This modular integration ensures a smooth and flexible coupling between Python-based optimization logic and TRNSYS simulations, allowing for real-time adjustments and rapid iteration across design scenarios.

In this study, each design configuration is evaluated using a deterministic simulation approach implemented in TRNSYS. As TRNSYS does not introduce stochastic variation in its outputs and each configuration is simulated only once, the results are consistent and repeatable. Consequently, statistical uncertainty does not arise, and the use of error bars or confidence intervals is not applicable.

2.2.4. Refinement Found (RF) and Early Stop

Each iteration of the optimization loop passes through the RF module. Here, the newly obtained data is assessed to identify any significant improvements or refinements over the previous iteration in the objective

functions. If refinement is found (RF is evaluated to *True*), there is no need to change the logic, and based on the constant data flow from the SM, the same strategy is applied in the DDDM module. If there is no refinement found (RF is evaluated to *False*), the Early Stop module is triggered to check if conditions are satisfied.

2.2.5. Change Logic (CL)

Informed by the DDDM and SM output, the CL module employs a range of techniques to effectively explore the search space. The incorporated strategies are summarized in Table 5.

It is important to note that there is no specific order in which a particular sampling method should be called. The choice is determined by the outcome received from the SM and the history provided by the DDDM module.

2.2.6. Change Significance Threshold (CST)

Following the CL module, the CST evaluates whether the changes observed warrant an adjustment in the significance level of variable alterations. This dynamic calibration is imperative for maintaining the optimization process's focus and adaptability.

2.2.7. Loop continuation and termination

The loop persists with iterations of data generation and refinement checks, becoming smarter and making more accurate predictions in terms of the feature importance of independent variables after each iteration, until specific stopping conditions (Early Stop) are met, such as a predefined tolerance level between the objective functions of consecutive iterations.

2.2.8. Anchor points refinement and TOPSIS

In multi-objective optimization, the resolution of conflicting objectives yields a set of trade-off solutions that balance economic and environmental goals. Traditional optimization methods, such as point-by-point search, may result in suboptimal results. To address this issue, the Pareto dominance criterion is used to identify the most efficient (non-dominant) solutions and eliminate those that are dominated by others. Before applying Pareto dominance, the refinement of anchor points is conducted to enhance the precision of the optimization process. The decision-making process is further refined using the Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) [64]. This method, when combined with variable weights, helps to select the final Pareto optimal solution that best balances the conflicting objectives.

Table 5
Sampling strategies applied in optimization.

Technique	Sampling type	Description	Reference
Latin hypercube sampling (LHS)	Stratified	Ensures uniform coverage across multidimensional parameter space, facilitating comprehensive modeling and simulation precision.	[35,36]
Focused random walks (FRW)	Heuristic	Also known as DDAH-SO*, generates samples around the best-known solution using Euclidean and Manhattan metrics to efficiently enhance the search.	[37–39]
Gaussian sampling around point (GASP)	Stochastic	Creates Gaussian-distributed samples around the current optimal point to explore the solution space with controlled randomness.	[40]
Particle swarm optimization (PSO)	Metaheuristic	Simulates the social behavior of particles to escape local optima and perform a global search, improving the optimization process.	[33,41]

* Dual-distance adaptive hyperparameter sampling for optimization.

2.2.9. Visualization and convergence

The optimization process concludes with the visualization of the Pareto frontier, which illustrates the trade-offs between environmental and economic goals. The convergence of the frontier indicates the effectiveness of the framework in identifying the most sustainable and cost-efficient design options for the energy system under consideration.

In summary, the proposed approach utilizes a data-driven decision-making module that serves as a strategic guide through the solution space, utilizing FIS within a machine-learning framework to quickly identify and fine-tune key design parameters. This systematic procedure significantly enhances the efficiency of the optimization process, leading to an optimal solution while minimizing resource requirements.

The constructed framework assesses the SDES across two primary metrics: economic performance, focusing on financial metrics over the system's lifespan, and environmental performance, evaluating the system's impact throughout its entire life cycle.

The proposed optimization framework relies on a tightly coupled TRNSYS-Python integration, where each environment fulfills distinct and complementary roles. Python orchestrates the entire optimization logic, including the initial design samples generation, adaptive selection of sampling strategies (statistical, heuristic, stochastic, metaheuristic), decision-making based on model feedback, and refinement routines. TRNSYS, in turn, serves as the high-fidelity simulation engine responsible for evaluating the energy performance of each system design configuration. During each iteration, Python generates input parameter files, triggers TRNSYS runs through automated scripting, and retrieves simulation outputs for post-processing and performance evaluation. These outputs are then used to update internal model performance metrics, guide subsequent decision-making, and determine convergence criteria. This modular integration ensures computational flexibility while avoiding surrogate approximations, allowing real-time adaptation to system behavior and enabling scalable exploration of complex multi-objective energy design spaces.

It should be emphasized that the proffered methodology's adaptability allows for the various algorithms to the CL and DDDM modules, enhancing the model based on the specific problem's requirements.

The exhaustive list of libraries used in optimization is presented in the annex, Table 1.

3. Case study

3.1. Climate and demand profiles

The small residential community of Falset (41°08′46″N 0°49′12″E), located in the Tarragona province of Spain, was chosen to apply the proposed SDES concept. The energy community comprises 1058 buildings with 1712 individual dwellings [65]. The proposed system has been developed to meet the requirements for Space Heating (SH) at 50 °C, Domestic Hot Water (DHW) at 60 °C [21], and to fulfill the electricity demand as well.

To compute the SH demand, the number of residential buildings' cluster types within the Falset municipality was obtained using cadastral data [66]. The energy demand per 100 square meters for each cluster type in Tarragona province has been obtained from [67]. The hourly SH demand was calculated by assessing the temperature difference between the hourly air temperature and a set temperature of 18 °C during the heating season of a typical year (from October to May), as per the methodology outlined in [68]. The utilized meteorological data was a location-specific TMY file downloaded from the Photo-voltaic Geographical Information System (PVGIS) platform [69]. The hourly SH demand was calculated by apportioning the annual total based on the relative difference in temperature for each hour.

The hourly fractional DHW demand in the residential sector was

calculated following the [70] methodology and shown in Table 6. The daily/hourly municipal tap water temperatures T_w, were computed according to [71].

The daily DHW energy demand $Q_{\mathrm{DHW}}(J)$ for the municipality is defined as:

$$Q_{\mathrm{DHW}} = Pop \bullet V \bullet \rho \bullet C \bullet \left(T_{\mathrm{sup}} - T_{\mathrm{w}}\right), \tag{16}$$

where:

Pop – population per municipality (number of people).
V – the DHW volume per person per day recommended by [72], 0.030 m^3.
ρ – water density, kg/m^3.
C – specific water heat capacity, J/kg °C.
T_{sup} – hot water supply temperature, °C [73].
T_{w} – tap water temperature, °C.

The residential electricity demand for Falset was based on the reported consumption for 2019 as obtained from the Datadis platform [74]. The total annual SH, DHW, and electricity demands for the residential buildings in Falset municipality are 10.61 MWh, 1.41 MWh, and 3.71 MWh respectively. The monthly values for the climatic data alongside the SH, DHW, and electricity demands obtained from the hourly data are shown in Fig. 2.

3.2. SDES specification and operation

To analyze and simulate the behavior of the SDES system, TRNSYS 18 software was used. This software uses partial differential equations and imposes constraints on mass and energy conservation under specified boundary conditions. The system's dynamic characteristics were well-captured, thanks to the software's ability to simulate the SDES realistically. For computational efficiency, the SDES model was simulated for three consecutive years. The results of these simulations were then extended across the project's entire lifespan, which was assumed to be 40 years, based on the assumption that weather conditions and demand profiles would remain consistent during this period. The SDES model validation was conducted based on the work of Abokersh et al. [53], Tulus et al. [18], and the original concept by Guadalfajara et al. [17]. The main SDES scheme is shown in Fig. 3.

3.2.1. Main SDES components and control strategies

The TRNSYS model incorporates a variety of components, each with specific parameters and input-output variables. The main components of the system are PV (type 190), STC (type 1a), Water-to-Water HP (type 927), Fully Stratified Storage Tanks (type 4c), Counterflow HEs (type 5b), and AUXs (type 6).

The SDES is aimed to meet the daily needs for SH, DHW, and electricity as shown in Fig. 3. This system mainly uses roof-mounted STC and PV, an SST, DHWT, a central water-to-water HP unit, and AUXs fueled by natural gas. The HP serves as a backup heat source for both SST and DHWT. In this setup, heat collected via the STC is used to meet SH/DHW demands, where it is transferred across the distribution network through HEs using 'Y' valves. Depending on the operational mode, the heat produced by the HP is either used for SH or DHW, with the potential to charge the SST.

During winter, the SST is used to meet SH demands, while the DHWT meets the daily DHW needs. SH is supplied at a low temperature (50 °C), whereas DHW is supplied at a higher temperature (60 °C). If solar collectors, HP, and SST cannot satisfy the heat load, AUXs cover the shortfall. Additionally, electricity is provided by a combination of on-grid roof-mounted PV panels and the electricity grid, meeting the neighborhood's electric demand and SDHS electric equipment.

An efficient control strategy is used to meet residential heating demands, maximizing the use of solar energy. Four operational modes are planned, based on the temperature levels of the SDHS, and are enabled via on/off control switches:

1) **Mode 1**: In the first mode, the heat from solar collectors is transported to the DHWT with the help of P_1, P_2, and P_5 pumps through HE_2. If solar thermal energy is insufficient for the DHW network, the auxiliary heater (AUX_2) is enabled.
2) **Mode 2**: In the second mode, SH is initiated when a suitable temperature in DHWT (T_{DWT}) is reached, and the temperature of the collector (T_{COL}) is higher than the bottom of SST (T_{SST}). P_1, P_2, and P_3 pumps transfer heat to SST from STC via HE_1.
3) **Mode 3**: In the third mode, both DHW and SH circuits are initiated when the criteria for both operations are met, where $T_{SST} > T_{DHW}$.
4) **Mode 4**: In the fourth mode, HP operates if the solar thermal system cannot meet heating demands. Here $T_{COL} < T_{SST}$, which is less than the reference turn-on temperature of the heat pump (T_{ref}). The heat generated by the HP is used for either SST or DHWT. Any uncovered heat demands are met by using AUXs.

Additionally, the on-grid PV panels reduce purchased electricity from the supply grid, covering both neighborhood electric demand and SDHS electric equipment.

3.2.2. Simulation model

The information flow diagram is shown in Fig. 4 (Type – inside TRNSYS GUI). Each component in the model is accompanied by information boxes detailing specific parameters and input-output variables.

The thermal behavior of fluid-filled sensible energy storage tanks (Type 4c), which experience thermal stratification, is captured by segmenting the tank into N ($N \leq 100$) fully mixed, equal-volume segments. The value of N indexes stratification's extent. At N equal to 1, the tank operates as a fully mixed unit, precluding any stratification effects. For this study, the tanks are segmented into 12 nodes to adequately represent stratification effects. In this configuration of Type 4, the stratified tank features variable inlet positions, allowing incoming fluid to be introduced at a temperature closely matching its own. Additionally, this setup presumes uniform loss across each tank node and does not account for loss to the gas flue of the AUX.

The model's prime components are listed in Table 7.

3.3. Economic and environmental input data

The parameters for the LCC assessment are outlined in Table 8. Following [18], the maintenance cost is estimated to be 1.5 % of the initial purchase cost of the equipment. The prices of natural gas and electricity are 0.0862 and 0.2871 €/kWh, respectively, based on the EUROSTAT [75] and Red Electrica [76] databases, and the price of electricity injected into the grid is 0.10 €/kWh [77]. Moreover, the inflation rates associated with natural gas and electricity are 5.6 % and 3.9 %, respectively [75,78]. The inflation rate associated with the proposed system throughout its lifetime was set to be 1.61 %, with a discount rate of 0.64 % [79,80].

The input parameters for the LCA were extracted from the Ecoinvent

Table 6
The hourly fractional DHW demand.*

Hour	f_h^*	Hour	f_h^*	Hour	f_h^*	Hour	f_h^*
0 h	0.01	6 h	0.03	12 h	0.05	18 h	0.05
1 h	0.00	7 h	0.10	13 h	0.05	19 h	0.07
2 h	0.00	8 h	0.07	14 h	0.04	20 h	0.06
3 h	0.00	9 h	0.07	15 h	0.03	21 h	0.06
4 h	0.00	10 h	0.06	16 h	0.04	22 h	0.05
5 h	0.01	11 h	0.06	17 h	0.04	23 h	0.05

* f_h refers to the fraction concerning the daily total DHW demand.

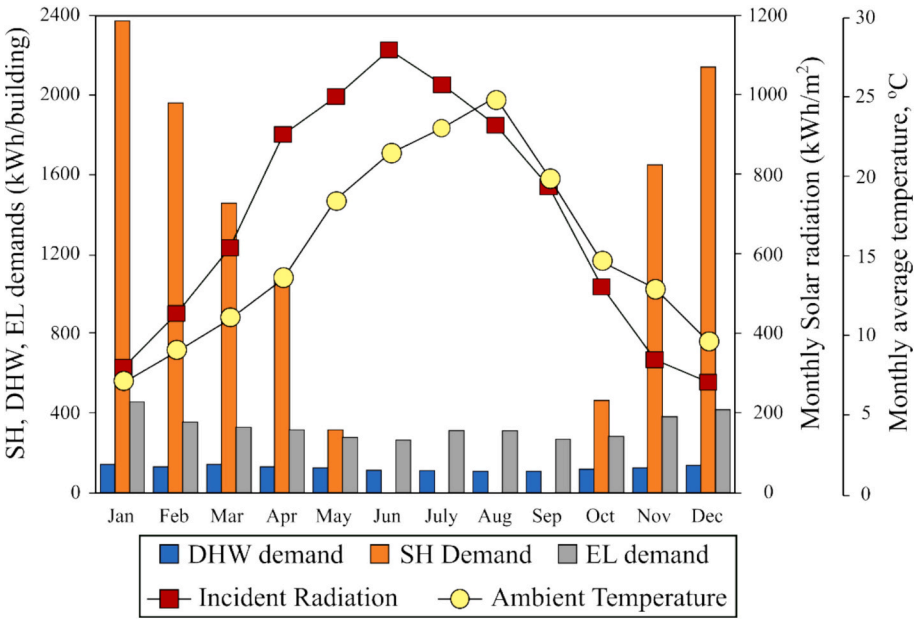


Fig. 2. The monthly climate conditions and the SH, DHW, and Electricity demand profile of the investigated energy community of Falset.

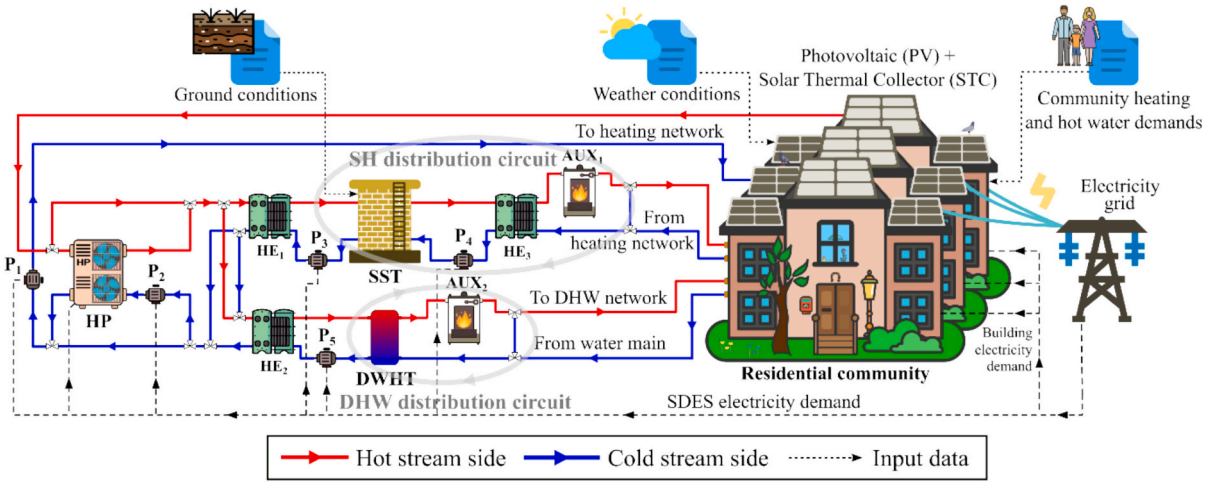


Fig. 3. Solar District Energy System (SDES).

v. 3.9.1 database [59]. It categorizes the damage associated with the SDES components based on the Recipe 2016 approach. The cumulative endpoint environmental impact is calculated with Eq. 2.11 and the specific impact of different components of the SDES is shown in Table 9.

3.4. Assumptions and system boundaries

This study operates under several foundational assumptions to guide the simulation and optimization of the SDES. It is assumed that the project has a lifespan of 40 years, during which key components such as STCs, PV panels, HP, HEs, DHWT, AUXs, and centrifugal pumps require replacement after 20 years of continuous operation [81] whereas the SST lifetime is set to 80 years [82]. The horizon of the dynamic simulation in TRNSYS is 3 years, with the 3rd year being extrapolated to represent the system's performance over the remainder of the project's lifespan. The time resolution for the TRNSYS simulations is set at 1 h. The rooftop area available for solar installations is uniformly distributed across all buildings, with solar collectors prioritized for placement, followed by PV panels. The available rooftop area is calculated following the Saez et al. [83] methodology. Any excess PV panels are installed in nearby fields due to space constraints. Climate data, derived from the PVGIS [69] platform as a TMY file, is assumed to remain constant throughout the project's duration. To optimize computational resources, the SDES model is simulated for three benchmark years, with the third year – representing a steady state – extrapolated over the remaining lifespan. The extended database includes decision variables (discrete values) such as the STC area, PV area, SST, and DHWT volumes, which are incremented in standard sizes due to practical construction limitations. The other prime assumptions for the model are outlined in Table 10.

Table 11 summarizes the lower and upper bounds defined for each design variable used in the optimization process, reflecting the feasible solution space explored by the applied sampling techniques.

The study considers only the thermal energy generated by the system for use within the connected energy community, without external sales. However, it posits unlimited potential revenue from selling any excess electricity generated to the grid, enhancing the system's economic viability. No negative environmental impact is considered; all the impact of electricity surplus sent to the grid is nullified; no allocation is applied. These assumptions are necessary to simplify the model while

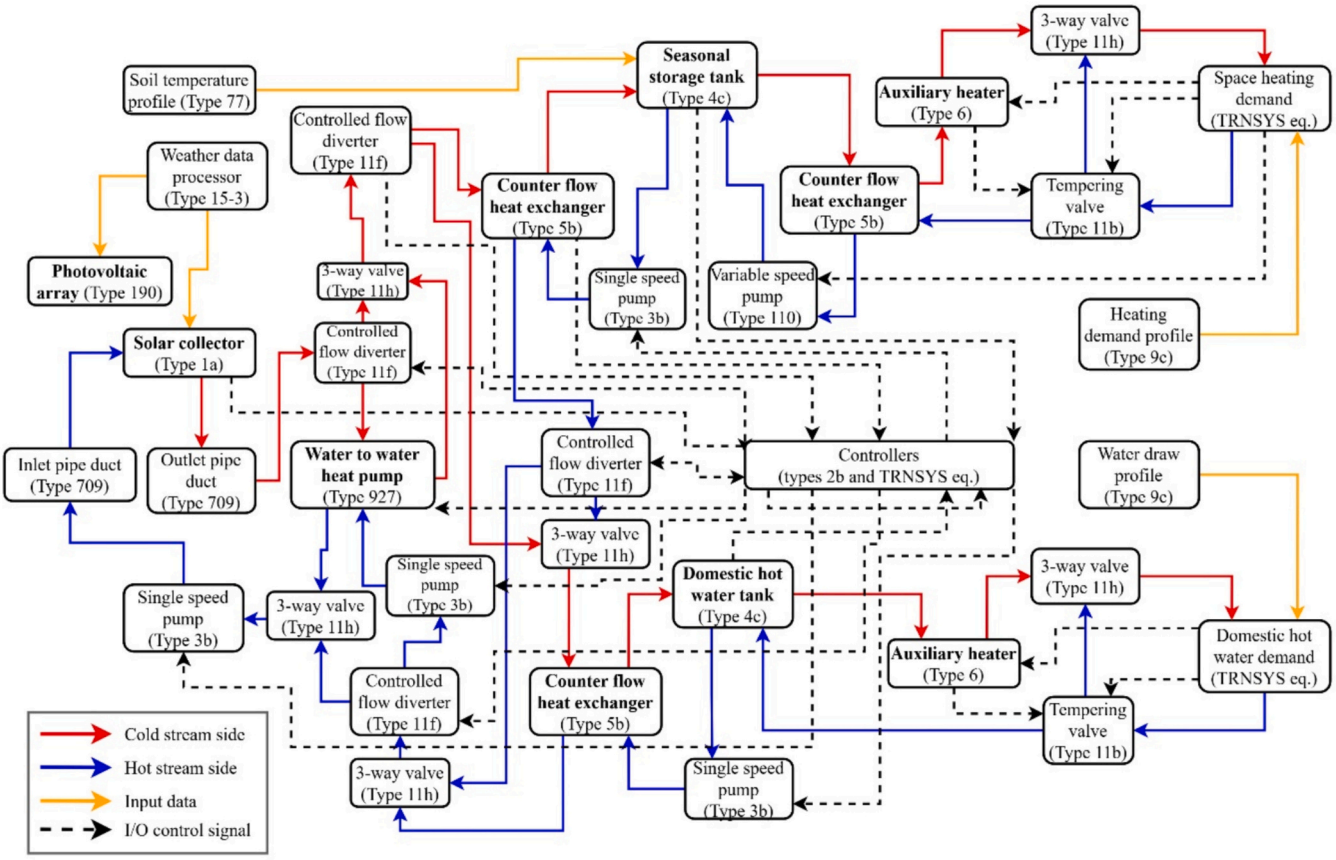


Fig. 4. Flow diagram of the SDES system simulated, with the components and their interconnections using TRNSYS 18.

ensuring its relevance and applicability to real-world scenarios, as well as allowing the full potential of the system to be revealed.

4. Results and discussion

This section presents the results of the proposed methodology. The model exhibits notable computational efficiency, averaging 50–70 CPU seconds per point when runs are executed sequentially on a 12th Gen Intel(R) Core(TM) i7-1255U processor with 16 GB RAM. This range was derived from actual runtime logs recorded during the optimization process, where the execution time of each TRNSYS simulation and subsequent Python post-processing was tracked. Occasional outliers due to system latency or background processes were excluded from the estimate. The reported average serves as a practical performance reference and may vary depending on the specific hardware configuration and system complexity. Python's versatility enables multiple concurrent TRNSYS simulations, with the maximum number only limited by the available computational resources, specifically the CPU cores. In this setup, it is possible to execute up to 12 concurrent runs simultaneously (a single executive run per processing core), with optimization averaging between 10,350 and 13,000 CPU seconds compared to 78,400 CPU seconds [21].

4.1. Comparative discussion with traditional optimization frameworks

Beyond its demonstrated computational efficiency, the proposed TRNSYS – Python hybrid framework delivers substantial methodological advantages over widely used alternatives, such as surrogate-based optimization (SM) and TRNSYS-GenOpt coupling. While surrogate models offer reduced computation time through approximation, they inherently compromise modeling accuracy. These approaches require extensive upfront data generation and frequent retraining when system parameters change, which not only increases the preparatory workload but also reduces adaptability in iterative design scenarios [31,32]. Moreover, surrogate models are prone to extrapolation errors when used outside the trained dataset, making them unsuitable for dynamic or evolving system configurations.

TRNSYS-GenOpt integration, on the other hand, supports several global optimization algorithms through a structured XML-based interface but suffers from limited flexibility. GenOpt does not allow dynamic reconfiguration of the optimization strategy during execution, meaning only one algorithm can be applied per run. Moreover, GenOpt lacks built-in data handling or postprocessing capabilities, often requiring manual intervention or scripting outside the optimization loop to process results, perform sensitivity analysis, or adjust decision variables between iterations [30]. This static coupling restricts its adaptability to complex multi-stage problems or the inclusion of decision-feedback loops.

In contrast, the proposed methodology enables a modular and dynamic optimization process where different sampling strategies – including statistical (LHS), metaheuristic (PSO), heuristic (FRW), and stochastic (GASP) – can be triggered based on the evolving performance of the system. The TRNSYS-Python integration supports seamless simulation execution, real-time refinement, adaptive sensitivity analysis through Feature Importance Scoring (FIS), and parallel processing for scalability. Unlike surrogate-based frameworks, it maintains full simulation fidelity throughout the optimization process, and unlike GenOpt, it provides algorithmic flexibility, automatic refinement logic, and robust data handling through Python scripting.

This flexible architecture ensures a more intelligent, adaptable, and high-fidelity search process, suitable for real-world SDES deployment scenarios.

While the focus of this study is not on benchmarking absolute statistical superiority over conventional methods, the methodological

Table 7
SDES' main components.

Component	Type	Description	Key Specifications
STC	Type 1a	Solar collectors for capturing thermal energy.	Optical efficiency: 0.817, heat loss coefficient: 2.205 W/m^2·K
PV	Type 190	Photovoltaic panels for converting solar energy to electricity.	Conversion efficiency: 0.186
Stratified storage tanks	Type 4c	Fully stratified tanks for storing heated water.	DHWT heat loss coefficient: 0.3125 W/m^2·K, SST heat loss coefficient depends on construction material
Water-to-water HP	Type 927	Heat pumps use water sources for heating and cooling.	Data from WSHP-PRC026G-TRANE, Heat capacity and power consumption across various temperatures
Counterflow HEs	Type 5b	Heat exchangers for transferring thermal energy between fluids.	Heat transfer coefficient: 3.931 W/m^2·K
Centrifugal pumps	Type 3b	Single-speed pumps for circulating water.	–
AUXs	Type 6	Backup heaters to supplement the solar thermal system.	Efficiency: 93 %
Inlet and outlet pipe ducts	Type 709	Ducts for directing flow into and out of system components.	–
Three-way valves	Type 11h	Valves to control the direction of flow in the system.	–
Tempering valves	Type 11b	Valves for mixing hot and cold water to achieve desired temperatures.	–
Controlled flow diverters	Type 11f	Devices for managing flow rates within the system.	–
Controllers	Type 2b	Devices to control the operation of various system components.	–
Soil temperature profiles for SST	Type 77	Soil temperature data to inform SST heat loss calculations.	–
Weather data	Type 15–3	External weather data for modeling system performance under varying conditions.	–
Time-dependent forcing functions	Type 9c	Profiles for heating and DHW demand over time.	–

reliability of the proposed framework is supported by several qualitative and quantitative indicators. Specifically, the framework maintains consistently low prediction errors (e.g., MSE), demonstrates stable generalization behavior between training and testing phases, and produces well-distributed Pareto fronts with clearly refined trade-off boundaries and anchor points. Although traditional confidence intervals or statistical significance tests are not applicable due to the deterministic nature of TRNSYS simulations and one-time evaluation of

Table 9
The environmental impact of SDES equipment based on ReCiPe 2016 [59].

Equipment/material	Options	Impact factor (ReCiPe 2016)	Unit
STC		10.2	Pt/m2
PV		15.7	Pt/m2
DHWT		43.0	Pt/m3
SST insulation	XPS	0.889	Pt/kg
	MW	1.88·10–3	Pt/kg
	FG	0.164	Pt/kg
SST construction	NC	0.01	Pt/kg
	HPC	2.2·10–3	Pt/kg
	UHPC	0.015	Pt/kg
HP		7.1	Pt/kW
HE		6.65·10–2	Pt/m2
Water pump		4.54	Pt/kW
AUX		4.1	Pt/kW
Natural gas		2.6·10–3	Pt/kWh
Electricity		0.0149	Pt/kWh

Table 10
Prime optimization scenario.

	Thermal energy use	Electricity selling policy	Environmental impact	Allocation*
Optimization scenario	All for community use	Expenses covered + additional revenue allowed	No negative impact considered; surplus impact nullified	Not considered

* Allocation: in energy systems, the process of distributing and assessing the impact of produced energy, focusing on the self-consumption share within a community and excluding surplus energy sent to the grid.

Table 11
Design variable boundaries used in the optimization procedure.

Decision variable	PV area, m^2	STC area, m^2	SST volume, m^3	DHWT volume, m^3
Optimization boundaries (from-to)	5000-30,500	4000-36,000	8000-140,000	210–605

Table 8
The economic parameters of SDES equipment [18].

Equipment, material	Options	α_k	β_k	CAP_k	Unit	Base year	BMF_k
STC		974.2	0.8330	Aperture area	m^2	2007	1
PV		190.92	1	Aperture area	m^2	2022	1
DHWT		11,680	−0.5545	Volume	m^3	2018	1
SST construction	NC	4178.1	−0.394	Volume	m^3	2000	1
	HPC	2575	−0.363	Volume	m^3	2004	1
	UHPC	90.83	−3	Volume	m^3	2004	1
SST insulation	XPS	561.09	0.3974	Material thickness	m	2017	1
	MW	1902.7	0.942	Material thickness	m	2017	1
	FG	311.41	0.9698	Material thickness	m	2014	1
HP		2053.8	−0.3480	Duty	kW	2014	1.4
HE		3.133	−0.331	Exchange area	m^2	2001	3.29
Pump (P_1, P_2, P_4, P_5)		389.0	283.15	Mass flow rate	kg/h	2009	3.24
Pump (P_3)		389.0	717.0	Mass flow rate	kg/h	2009	3.24
AUX		225.0	0.7460	Duty	kW	2001	2.1

design points, convergence consistency and cross-validation performance serve as practical proxies for robustness. Together, these aspects affirm the credibility and practical optimization reliability of the proposed method in comparison with more static, monolithic approaches.

4.2. Optimization metrics and model performance evaluation

4.2.1. Model configuration

In constructing the predictive model and subsequent decision-making analysis, a meticulously balanced approach was employed to account for both environmental and economic factors. The TOPSIS was applied, where environmental impact and economic cost were weighted across a spectrum from exclusivity to another, as detailed in Table 12.

For predictive accuracy, the Gradient Boosting Regressor (GBR) was fine-tuned with hyperparameters including a learning rate of 0.01 and max depth of 2, among others listed in Table 13.

To ensure the model's robustness and guard against overfitting, a k-fold cross-validation approach was adopted with shuffled data splits, enhancing the reliability of the model predictions (Table 14).

Adding shuffling and noise helps to ensure that the cross-validation process not only evaluates the model thoroughly across different subsets of data but also assesses its ability to perform well in practical, real-world scenarios where data may not be perfect. These methodological choices set the groundwork for the ensuing results and discussion, providing a robust analytical framework for interpreting the model's performance.

4.2.2. Model performance and convergence evaluation

In optimizing complex systems such as SDES, the accuracy with which a model predicts the impact of various system components – termed 'prediction accuracy' – is pivotal. This metric evaluates how closely the model's forecasts align with actual outcomes, serving as a critical indicator of the model's reliability and effectiveness. High prediction accuracy is crucial for prioritizing input decision-making variables, enabling the model to quickly identify and adjust the most impactful variables to approach a global optimum efficiently. Conversely, exploring less significant variables, while less critical for reaching the global optimum, is valuable for fine-tuning and local search, ensuring thorough optimization. This section delves into the model's capability to distinguish between these variables with precision, demonstrating its proficiency in guiding strategic modifications that enhance both the environmental and economic efficacy of the SDES.

To assess the predictive accuracy of the model, MSE is among the most commonly used or recommended measures [84]. For this reason, the Mean Square Error (MSE) is considered in this study.

In regression analysis, feature importance scoring helps us understand which independent variables are most influential in predicting a dependent variable. Essentially, the process involves feeding a set of data that includes both independent and dependent variables into a regression model. The model then performs calculations to quantify the importance of each independent variable in terms of a percentage, representing its contribution to predicting the outcome. All these percentages add up to 100 %, showing the relative importance of each feature.

Mean Squared Error (MSE) is a measure used to evaluate the accuracy of the model's predictions. It calculates the average squared difference between the actual and the predicted objective functions by the model. MSE is computed separately for the training data (which the model has seen) and the testing data (which it hasn't seen). Comparing train and test MSEs helps to assess how well the model performs in predicting new, unseen data and, thus, how reliable the feature importance scores are. It also serves as a convergence analysis, demonstrating the model's ability to generalize effectively without overfitting.

While this method uses the model to predict outcomes as a step in calculating feature importance, it's not a surrogate model in the strictest sense. Surrogate models typically replace more complex models entirely, but here, the prediction is just an intermediate step used to understand feature influence before the final scoring of feature importance is derived by regression. This approach gives us a straightforward, practical insight into which factors are most significant in influencing the predicted outcomes.

The obtained MSEs (Fig. 5) indicate the model's high accuracy and reliable performance in the feature importance prediction in RCP and NPC optimization. The MSE values for training and testing sets are consistently low across all features, demonstrating minimal prediction errors. For NPC optimization, the maximum difference in MSE is observed in the STC area with a train MSE of 0.04213 and a test MSE of 0.04261, resulting in a 1.13 % error. The minimum difference is registered in the PV area, with a train MSE of 0.05332 and a test MSE of 0.05354, reflecting a slightly lower error of 0.41 %. In RCP optimization, the maximum difference is found in the SST volume, where the train MSE is 0.03352 and the test MSE is 0.03426, showing a 2.16 % error. Conversely, the minimum difference is in the PV area, with train and test MSE values of 0.04013 and 0.04045, respectively, leading to a 0.79 % error. These results highlight the model's robust performance and effective generalization, even in the features with the highest errors between training and testing sets to say nothing of the lowest errors.

Each Gradient Boosting Regressor (GBR) model is trained deterministically on a unique dataset generated during each optimization iteration. Since each design configuration is simulated only once via TRNSYS and no repeated samples exist, traditional error bars or confidence intervals are not applicable. Instead, model reliability is assessed through convergence between training and testing performance. The percentage differences shown above each pair of bars in Fig. 5 represent this convergence, serving as a practical indicator of generalization quality and ensuring the statistical robustness of the reported feature importance scores.

The importance ranking of each feature in both RCP and NPC optimization is the same, ranging from the most significant (PV area) with ranking I (maximum impact on the objective function) to the least significant (DHWT volume) with ranking IV (minimum impact on the objective function).

The numerical breakdown of feature importance, prediction errors, and ranking for RCP and NPC optimization is shown in Table 15.

4.3. Multi-objective optimal solutions

The SDES assessment affirms the model's reliability through environmental and economic performance evaluation. Fig. 6 displays this multi-faceted robustness, highlighting the environmental impact and economic cost in the anchor points A and C with various trade-off scenarios B1-B5 (system designs). It is a set of optimal solutions that define the Pareto frontier obtained from the optimization procedure.

The base case – depicting the system's environmental impact and economic cost without integrating renewable technologies – serves as a benchmark, set at 5.54 Pt/MWh and 336.0 €/MWh for the specific impact and highest specific economic cost (SEC) respectively. Point A, indicative of the optimal RCP, registers the lowest value at 3.709 Pt/MWh with the highest SEC of 108.54 €/MWh. Conversely, the optimal SEC solution (point C) has a minimum of 41.98 €/MWh with an RCP of 4.93 Pt/MWh. The trade-off points represent the intermediate non-dominant scenarios (points B_1-B_5). Moving from the best RCP to the

Table 12
TOPSIS weights for environmental and economic indicators.

Indicator	Applied weights						
	A	B_1	B_2	B_3	B_4	B_5	C
Environmental impact (RCP)	100 %	88 %	75.0 %	50.0 %	25.0 %	12.0 %	0 %
Economic cost (NPC)	0 %	12 %	25.0 %	50.0 %	75.0 %	88.0 %	100 %

Table 13
Hyperparameters for gradient boosting regressor model.

Applied regression	Gradient boosting regressor							
Model parameters	Random state	N° estimators	Learning rate	Max depth	Min samples split	Validation fraction	N° no change iterations	Tolerance
	0–99,999	50	0.01	2	2–10 %	0.1	10	0.0001

Table 14
k-Fold cross-validation configuration.

k-Fold cross-validation			
N folds (splits)	Shuffle	Noise	Random state
10	True	True	0–99,999

best SEC solution, the cost shows a notable reduction of 61.32 % whereas the environmental impact grows by 24.8 %. The RCP and SEC show that all optimal solutions have better environmental and economic performance than the base case scenario. The minimum environmental impact solution registered a 33.1 % drop when the minimum cost plummeted by an impressive 87.5 % as opposed to the base case.

The model considers multiple decision variables for different system designs to arrive at the optimal solutions described above and depicted in Fig. 6. The optimal design of the SDES also becomes an important consideration to achieve the desired economic and environmental performances. An in-depth analysis of the optimal solutions obtained provides a guideline for the SDES components sizing. There are four main decision variables: PV area, STC area, SST volume, and DHWT volume. These decision variables vary and provide a guideline for optimal system design. This range is shown in Fig. 7. Each diagram has lower and upper wicks (representing the optimization boundaries of each component), and a center solid body (representing the optimum size range of each component).

It should be pointed out that due to the applied restrictions (Section 3.4), the optimal area for STC achieves a simultaneous minimum in both NPC and RCP, resulting in a singular optimal point beyond which any expansion leads to increases in NPC and RCP, as surplus heat cannot be monetized through grid sales. This creates a scenario where the STC's optimal area is fixed, reflecting the balance between economic and environmental efficiencies under the given constraints. Such a unique phenomenon is referred to as reaching a utopian (idealized) solution [85].

Conversely, the PV area continues to vary (Fig. 7) because the revenue from selling excess electricity allows for a continued reduction in NPC even as environmental impact increases, creating a broader range of optimal solutions. This distinction reflects the inherent differences in how STC and PV technologies interact with the system's economic and environmental constraints, leading to a fixed optimal STC area while enabling flexible PV expansion for cost-effective and environmentally balanced solutions. The breakdown of optimal system component sizes is shown in Table 16.

It is noteworthy that the DHWT volume is ranked fourth on the Feature Importance Scoring (Table 15). Its variations result in minimal fluctuations in the objective functions, namely, environmental impact and cost. The variations in the objective function remain below 1 %, specifically within the range of 0.2–1 %. This minimal variation poses a significant challenge even for high-fidelity models to detect trends in the changes of this variable, as reflected in its variability (Table 16).

4.4. Environmental analysis

To support a detailed environmental analysis, Fig. 8 presents an in-depth breakdown of the environmental impact contribution of each parameter from A to C the Pareto optimal solutions over their operational lifetimes, alongside the base case solution. The numbers on top of each bar represent the overall (non-specific) environmental impact in ReCiPe points.

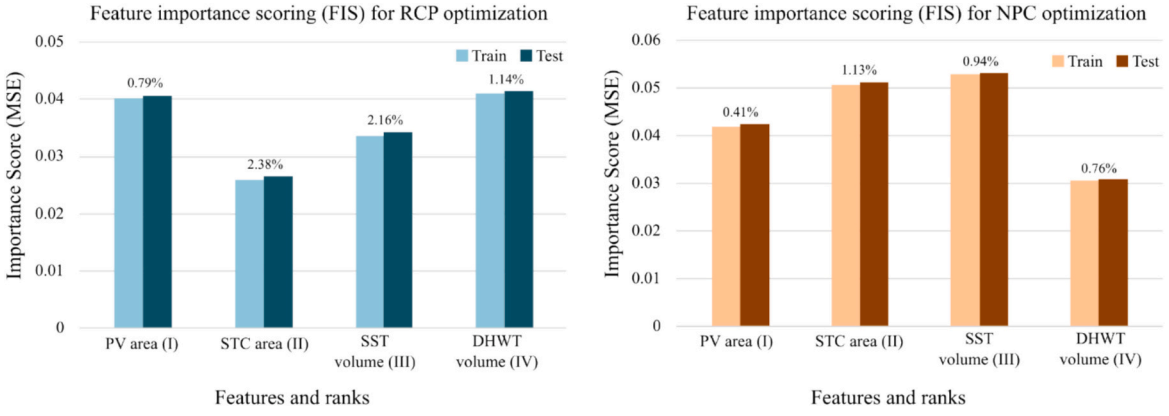


Fig. 5. Feature importance and convergence evaluation for RCP and NPC optimization.

Table 15
Detailed feature importance metrics for RCP and NPC optimization.

Feature	RCP optimization				NPC optimization			
	Train MSE	Test MSE	Prediction error	Rank	Train MSE	Test MSE	Prediction error	Rank
PV area	0.04013	0.04045	0.79 %	I	0.05332	0.05354	0.41 %	I
STC area	0.02582	0.02645	2.38 %	II	0.04213	0.04261	1.13 %	II
SST volume	0.03352	0.03426	2.16 %	III	0.03062	0.03091	0.94 %	III
DHWT volume	0.04088	0.04135	1.14 %	IV	0.05111	0.05150	0.76 %	IV

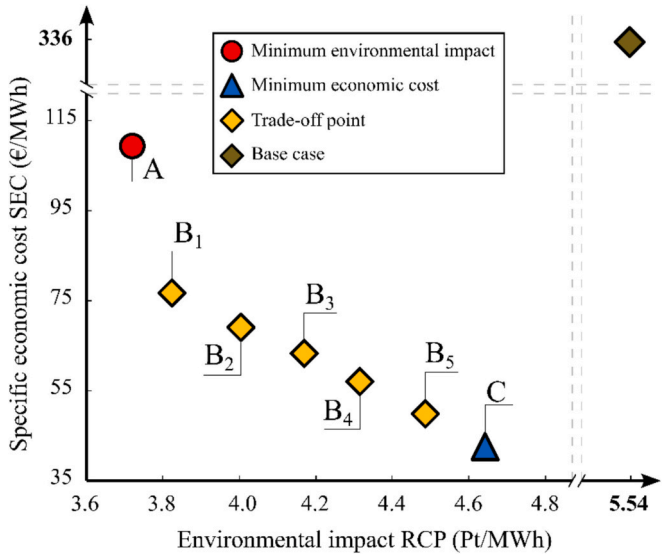


Fig. 6. Pareto set of optimal solutions for the SDES in Falset (A – min impact, C – min cost, B_1-B_5 – intermediate/trade-off points).

As figure shows, solutions A and C have reduced the environmental impact by deploying renewable technologies and optimizing the usage of non-renewable energy sources. The base case scenario serves as a benchmark with 3.76×10^6 environmental points and accounts for 100 %. The minimum cost solution (C) has a moderate 17 % reduction in environmental impact, while the minimum impact solution (A) shows a 32.7 % decrease in this regard.

In the base case, the natural gas and electricity impact account for 35.8 % (1.35×10^6 Pt) and 63.5 % (2.39×10^6 Pt) respectively while the AUXs impact has an inconsequential 0.7 % (0.02×10^6 Pt). Contrariwise, the natural gas contribution with renewable technologies applied shows a remarkable 97.8 % decline (0.03×10^6 Pt) in comparison to the base case. Electricity impact demonstrates a 33.2 % (1.60×10^6 Pt) and 34.8 % (1.53×10^6 Pt) drop for solutions A and C respectively.

The overall environmental impact trends and component-specific indicators are consistent with the findings reported in [53], demonstrating similar relationships between system configurations and environmental performance.

A complete breakdown of environmental analysis for the Pareto optimal solutions and the base case is provided in the annex, Table 2.

4.5. Economic analysis

As for the economic analysis, a detailed breakdown of the initial investment is shown in Fig. 9 below. The numbers at the top of each bar in Figs. 9–12 denote the total (non-specific) cost and are expressed in millions of euros.

As depicted in Fig. 9, the initial investment in the base case scenario is modest at €835,000, primarily driven by auxiliary heaters (€649,800 or 81 %) and water pumps (€47,100 or 5.9 %), with the remaining €138,200 (13 %) allocated to contingencies. In contrast, deploying renewable technologies leads to a 25- to 32-fold increase in total investment, ranging from €20.3 million (Solution A) to €25.6 million (Solution C). The seasonal storage tank (SST) is the largest contributor across all scenarios, reaching €8.01 million (35 % in Solution C), followed by photovoltaic panels (PV) at up to €5.33 million (21 %).

Solar thermal collectors (STC) and heat exchangers jointly contribute €3.46 million (14 %) and €2.93 million (11 %), respectively. The share

Table 16
Detailed SDES component sizes.

Component		STC, m^2	PV, m^2	SST, m^3	DHWT, m^3
Optimal solution					
Anchor point	A	10,818	9700	97,369	605
Parity point	B_3	10,818	19,300	108,188	506
Anchor point	C	10,818	28,900	129,826	605

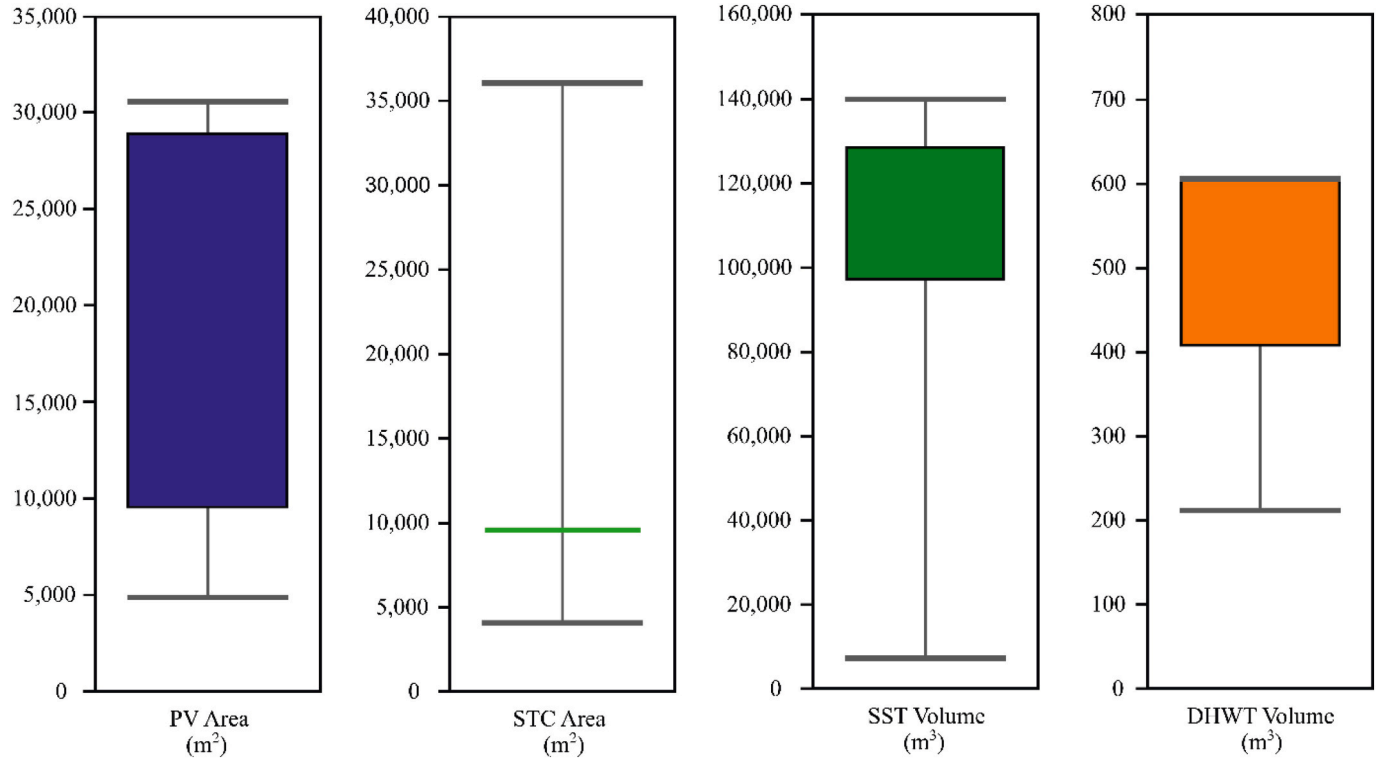


Fig. 7. Optimal component size ranges for SDES.

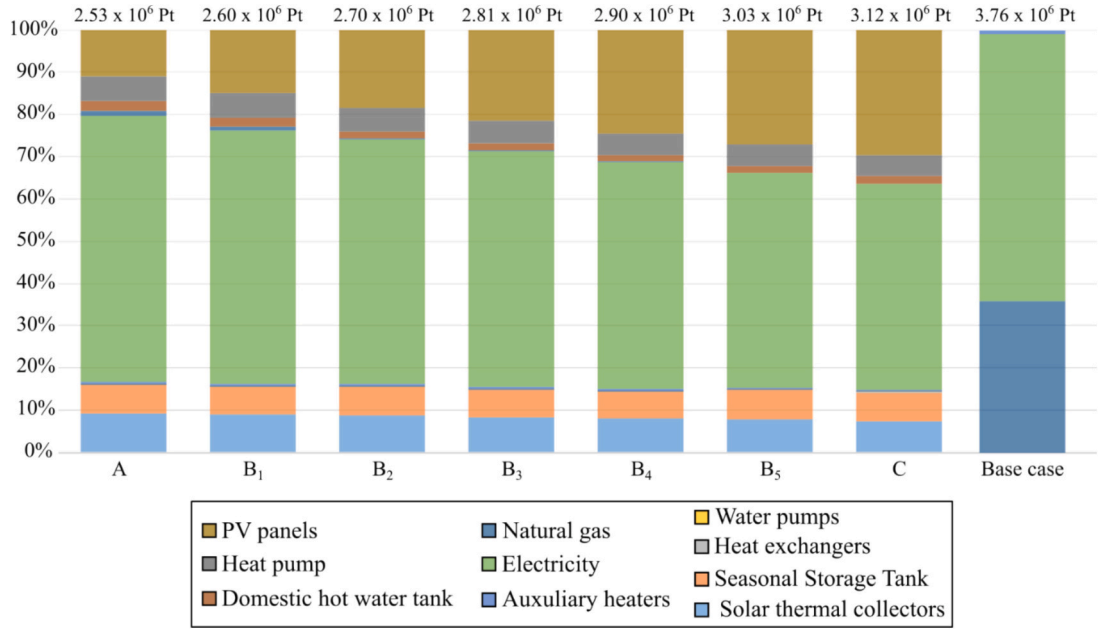


Fig. 8. Environmental analysis of optimal solutions in comparison to the base case.

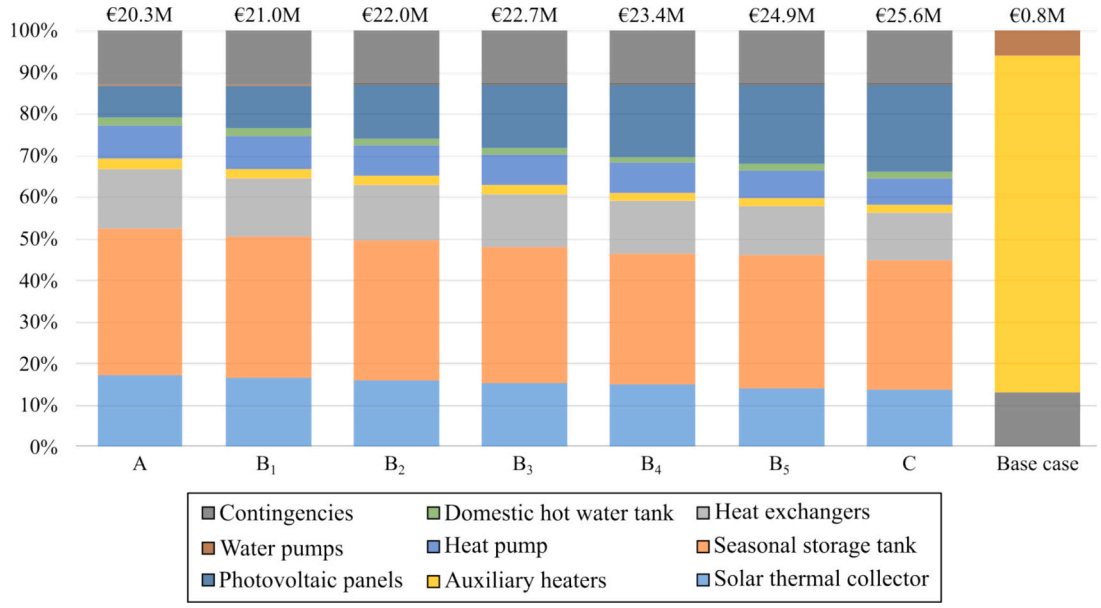


Fig. 9. Capital cost analysis of optimal solutions from the most sustainable to economically viable solution compared to the base case.

of auxiliary heaters drops significantly to around €487,000 (1.9 %), and the DHWT remains below €375,000 (1.5 %). Water pump costs are negligible at ~€50,000 (0.2 %). Notably, contingencies remain fixed at €3.34 million (13 %) across all optimal solutions. This cost breakdown highlights the sustainable scenarios' capital-intensive nature and investment drivers' shifting composition.

As shown in Fig. 10, the base case registers the highest operational expenditure, totaling €210.76 million over the 40-year system lifetime. The primary cost drivers are natural gas (€122.41 million, 58.1 %) and electricity (€88.10 million, 41.8 %), while maintenance accounts for just €0.25 million (0.1 %). Although the initial investment in this scenario is modest, the long-term reliance on fossil fuels and grid electricity results in low economic efficiency.

In contrast, renewable-integrated solutions drastically reduce operating expenses. Solution A, the most sustainable scenario, lowers total costs to €36.29 million – an 82.8 % reduction. This includes €10.71 million in maintenance, €22.62 million in electricity expenses, and just €2.67 million for natural gas.

Scenarios B_1 to B_3 represent a transition phase where electricity exports begin offsetting operational costs. In B_1, electricity revenues nearly match electricity expenses – approximately €82,000 in revenue – resulting in a total cost of €14.03 million, with €11.17 million in maintenance, €2.67 million in natural gas, and minimal net electricity-related expenditure. This scenario is notable for being the only one where electricity consumption and export are nearly balanced, meaning that while operating costs are still fully accounted for, electricity sales already provide a compensating effect. In B_2, net electricity revenue rises to €6.22 million, reducing total operational cost to €7.17 million, while in B_3, the impact of grid electricity sales becomes more prominent, with revenues reaching €11.78 million, bringing the overall operational

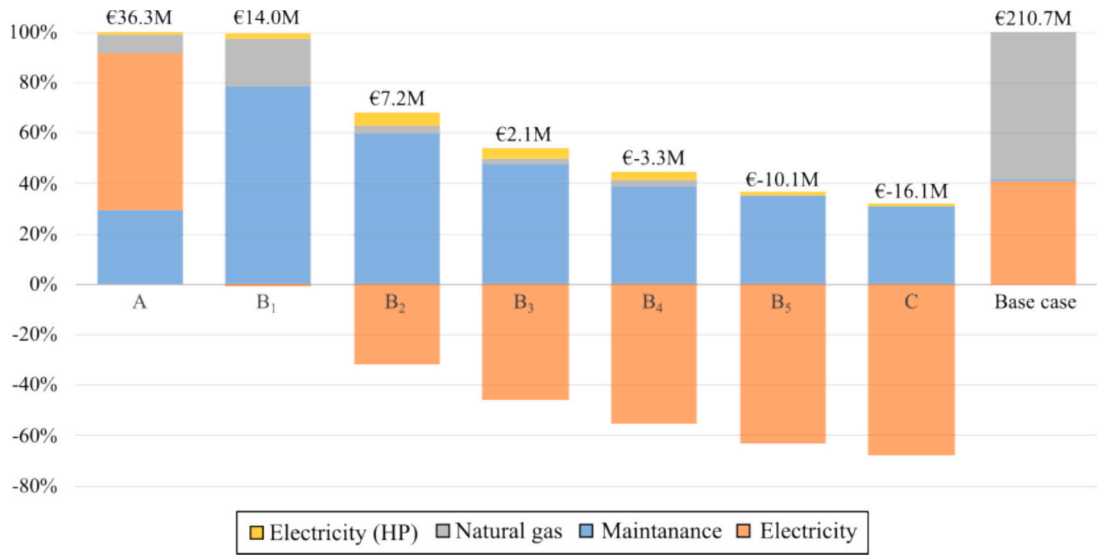


Fig. 10. Operational cost analysis of optimal solutions from the most sustainable to economically viable solution compared to the base case.

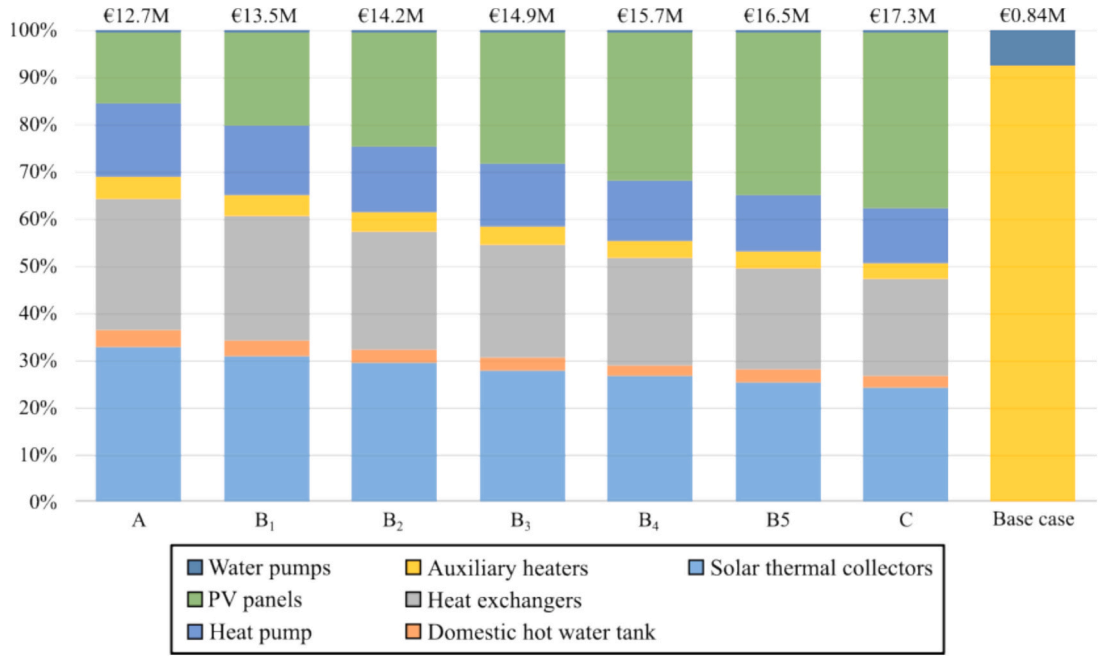


Fig. 11. Replacement cost analysis of optimal solutions from the most sustainable to economically viable solution compared to the base case.

cost down to €2.07 million.

Scenarios B_4 to C achieve full economic self-sufficiency, where electricity export revenue not only offsets expenses but also generates a net operating profit. In B_4, electricity sales reach €17.86 million, while total expenses – including €12.67 million in maintenance and €0.86 million in natural gas – are fully covered, resulting in a net surplus of nearly €3.3 million. B_5 further amplifies this trend, with €24.24 million in electricity sales, covering €13.62 million in maintenance, leaving an overall surplus of €10.1 million. Solution C offers the most favorable outcome, with €30.71 million in electricity revenue and €14.07 million in maintenance costs, resulting in a net gain of €16.15 million over the system's operational period.

These results highlight that renewable-integrated solutions not only reduce dependency on external energy sources but can also transform operational costs into long-term financial gains, reinforcing their viability under both economic and sustainability criteria.

Replacement costs exhibit similar trends to capital investments (Fig. 11), with minimal expenditures in the base case compared to renewable-integrated alternatives. The base case has a modest total replacement cost of €841,200, largely driven by AUXs (93 %, €781,300) and water pumps (7 %, €59,900). In contrast, solution A incurs a 15-fold increase, with total replacement costs of €12.72 million, reflecting the maintenance needs of additional components over the 40-year system lifetime.

In renewable-based solutions, AUX contributions remain below 5 %, while major contributors shift toward PV panels (up to 37 %), solar thermal collectors (STC, up to 33 %), and heat exchangers (HEs, up to 28 %). Heat pumps (HP) also present a significant share, reaching €1.99 million, or approximately 16 % of total costs in all scenarios. For instance, in solution C, the highest replacement cost scenario, PV replacement alone amounts to €6.45 million (37 %), followed by STCs (€4.18 million, 24 %), HEs (€3.55 million, 21 %), and HPs (€1.99

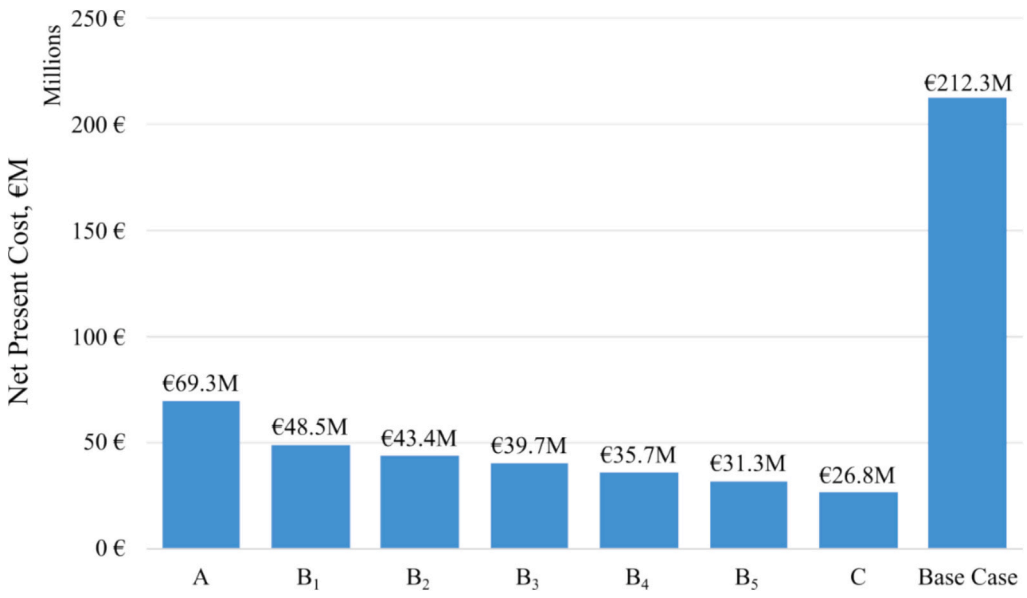


Fig. 12. Net present cost of the solar district energy system and the base case.

million, 12 %).

Intermediate scenarios B_1–B_5 show a gradual rise in overall costs from €13.48 million (B_1) to €16.51 million (B_5), driven mainly by incremental increases in PV capacity and marginal DHWT variations. Despite this, AUX replacements remain constant at €589,300, further emphasizing the reduced reliance on auxiliary heating units in renewable-based designs.

These results underline that although renewable technologies entail higher lifecycle replacement costs, the investment distribution shifts toward long-lived, high-efficiency components, significantly reducing dependence on fossil-based elements and enhancing long-term sustainability.

Finally, to conclude the discussion, Fig. 12 presents the Net Present Cost (NPC) analysis, offering a comprehensive assessment of the overall economic efficiency of each proposed solution across the 40-year system lifetime. The baseline scenario exhibits the highest total NPC, reaching €212.3 million, primarily driven by excessive operational expenditures (€210.8 million) due to heavy dependence on fossil fuels and grid electricity. In contrast, all renewable-integrated scenarios demonstrate substantial cost reductions and improved economic viability over the long term.

Among these, scenario A – the most environmentally sustainable configuration – achieves a 67 % reduction in total NPC, lowering it to €69.3 million, mainly due to a significant drop in operating costs despite moderately higher replacement investments. Notably, more balanced solutions along the sustainability-economics trade-off, such as scenarios B_3 and B_4, yield even more favorable economic outcomes, with total NPCs of €39.75 million and €35.74 million, respectively.

Scenario C, optimized for economic performance, reaches the lowest NPC at €26.76 million, reflecting an 88 % reduction compared to the baseline. This result is driven by the interplay of long-term investment efficiency, effective replacement cost planning, and consistent revenue streams from electricity export – reflected as net gains in operational cost.

These findings confirm that, although renewable configurations may entail greater upfront and replacement expenditures, their reduced dependency on fossil fuels and capability to generate financial returns make them superior in long-term economic terms. The NPC analysis serves as a final validation that the proposed hybrid optimization framework successfully identifies cost-effective, environmentally sound, and future-ready SDES configurations that far outperform the conventional fossil-based baseline system.

The total NPC of the SDES (Eq. (1)) alongside the base case is presented in the following figure.

The economic performance trends for the system components align with the outcomes observed in [53], following comparable cost-reduction patterns and efficiency improvements.

A detailed breakdown of the total net present cost alongside capital, operational, and replacement costs for the Pareto optimal solutions and the base case are provided in the annex, Tables 3, 4, 5, and 6.

5. Conclusions

The main objective of this study was to develop a new optimization methodology and prove its effectiveness through application to a real-world case study. This was achieved through a rigorous methodology that combines machine learning and advanced optimization techniques to optimize the Solar District Energy System (SDES) design. The Gradient Boosting Regressor model demonstrated robust performance with minimal prediction errors, ensuring reliable guidance throughout the optimization process. Unlike conventional approaches that heavily rely on heuristic, metaheuristic, or surrogate models, the proposed methodology uniquely combines heuristic rules with metaheuristic, statistical, and stochastic algorithms without generating extensive data. This blend thoroughly explores the solution space, effectively avoiding local minima and significantly reducing computation time.

The methodology was validated through a real-world case study in Falset, Spain, proving its practical utility and efficiency. It substantially reduced both environmental impact and economic cost (32.7 % and 87.4 % decline concerning scenarios A and C, respectively). Such an approach is particularly valuable for complex and dynamic energy systems where traditional optimization techniques are insufficient and only the parent model can be used.

In conclusion, the optimized SDES offers significant advantages over the base case, providing a financially and environmentally superior solution. By reducing the environmental footprint and economic cost, the optimized system supports the widespread adoption of renewable energy technologies in district energy systems. The findings of this study underscore the importance of advanced optimization techniques in achieving sustainable and economically viable energy solutions, offering a valuable framework for future developments in this field.

While uncertainty propagation was not within the scope of this study due to the deterministic nature of TRNSYS-based modeling, it remains an important and promising area for future research. In particular,

exploring how fluctuating economic or environmental parameters affect system performance could significantly enhance the robustness and real-world applicability of the proposed framework. Additionally, the scalability of the method under high-dimensional decision spaces will be systematically evaluated in future studies to test its resilience and adaptability. This includes stress-testing the optimization framework with an increased number of design variables, which is essential for modeling more complex energy systems. These enhancements will collectively contribute to the continued refinement and broader applicability of the optimization framework.

CRediT authorship contribution statement

Ruslan Kotegov: Visualization, Investigation, Writing – original draft, Software, Formal analysis, Methodology. **Mohamed Abokersh:** Methodology, Validation, Resources. **Carles Mateu:** Methodology, Writing – review & editing, Conceptualization. **Adedamola Shobo:** Writing – review & editing, Resources. **Dieter Boer:** Project administration, Funding acquisition, Conceptualization, Validation, Writing – review & editing, Data curation, Supervision. **Manel Vallès:** Validation, Project administration, Supervision, Funding acquisition, Conceptualization, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Manel Valles reports financial support was provided by Ministry of Science, Innovation and Universities. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge financial support from the “Ministry of Science, Innovation and Universities” of Spain (PID2021-127713OA-I00, PID2021-123511OB-C31, PID2021-123511OB-C33, PID2021-124139NB-C22 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE & TED2021-129851B-I00 funded by MCIN/EI/10.13039/01100011033 and by the “European Union NextGenerationEU/PRTR”) and the Maria Zambrano grant [2021URV-MZ-11] through “Ministry of Science, Innovation and Universities”, Spain - Recovery, Transformation and Resilience Plan; “The European Union – NextGenerationEU”; and “University of Rovira i Virgili”

Data availability

Data will be made available on request.

References

[1] E. Caramizaru, A. Uihlein, Energy Communities: An Overview of Energy and Social Innovation, Publications Office of the European Union, Luxembourg (Luxembourg), 2020 doi:10.2760/180576 (online).
[2] S. Bouckaert, A. Fernandez Pales, C. McGlade, U. Remme, B. Wanner, L. Varro, et al., Net Zero by 2050 – Analysis - IEA, International Energy Agency (IEA), 2021.
[3] Office of Energy Efficiency & Renewable Energy. Solar Futures Study | Department of Energy n.d. https://www.energy.gov/eere/solar/solar-futures-study (accessed October 9, 2023).
[4] M.K.H. Rabaia, M.A. Abdelkareem, E.T. Sayed, K. Elsaid, K.J. Chae, T. Wilberforce, et al., Environmental impacts of solar energy systems: a review, Sci. Total Environ. (2021) 754, https://doi.org/10.1016/j.scitotenv.2020.141989.
[5] A.L.G. Pires, P. Rotella Junior, L.C.S. Rocha, R.S. Peruchi, K. Janda, R. de C. Miranda, Environmental and financial multi-objective optimization: hybrid wind-photovoltaic generation with battery energy storage systems, J Energy Storage (2023) 66, https://doi.org/10.1016/j.est.2023.107425.
[6] S.B. Artaş, E. Kocaman, H.H. Bilgiç, H. Tutumlu, H. Yağlı, R. Yumrutaş, Why PV panels must be recycled at the end of their economic life span? A case study on recycling together with the global situation, Process Saf. Environ. Prot. 174 (2023) 63–78, https://doi.org/10.1016/j.psep.2023.03.053.
[7] P.A. Petroli, P.S.S. Camargo, R.A. de Souza, H.M. Veit, Assessment of toxicity tests for photovoltaic panels: a review, Curr Opin Green Sustain Chem (2024) 47, https://doi.org/10.1016/j.cogsc.2024.100885.
[8] J. Gao, H. Men, F. Guo, P. Liang, Y. Fan, A multi-criteria decision-making framework for the location of photovoltaic power coupling hydrogen storage projects, J Energy Storage 44 (2021) 103469.
[9] L.S. Paraschiv, S. Paraschiv, Contribution of renewable energy (hydro, wind, solar and biomass) to decarbonization and transformation of the electricity generation sector for sustainable development, Energy Rep. 9 (2023) 535–544, https://doi.org/10.1016/j.egyr.2023.07.024.
[10] J. Coughlin, A. Kandt, Solar Schools Assessment and Implementation Project: Financing Options for Solar Installations on K-12 Schools, 2011.
[11] M.H. Kim, D. Kim, J. Heo, D.W. Lee, Techno-economic analysis of hybrid renewable energy system with solar district heating for net zero energy community, Energy (2019) 187, https://doi.org/10.1016/j.energy.2019.115916.
[12] NASA, Solar Energy System Economic Evaluation Final Report for Wormser Columbia, South Carolina. United States, 1980, https://doi.org/10.2172/6509838.
[13] U. Aickelin, A. Clark, Heuristic optimisation, J. Oper. Res. Soc. 62 (2011) 251–252, https://doi.org/10.1057/jors.2010.160.
[14] A.A. Juan, P. Keenan, R. Martí, S. McGarraghy, J. Panadero, P. Carroll, et al., A review of the role of heuristics in stochastic optimisation: From metaheuristics to learnheuristics, Ann. Oper. Res. 320 (2023) 831–861, https://doi.org/10.1007/s10479-021-04142-9.
[15] M. Fischetti, M. Fischetti, Matheuristics, in: Handbook of Heuristics vol. 1–2, 2018, https://doi.org/10.1007/978-3-319-07124-4_14.
[16] N.E. Batista, P.C.M. Carvalho, L.M. Fernández-Ramírez, A.P.S. Braga, Optimizing methodologies of hybrid renewable energy systems powered reverse osmosis plants, Renew. Sustain. Energy Rev. (2023) 182, https://doi.org/10.1016/j.rser.2023.113377.
[17] M. De Guadalfajara, M.A. Lozano, L.M. Serra, Evaluation of the potential of large solar heating plants in Spain, Energy Procedia 30 (2012) 839–848, https://doi.org/10.1016/j.egypro.2012.11.095. Elsevier Ltd.
[18] V. Tulus, D. Boer, L.F. Cabeza, L. Jiménez, G. Guillén-Gosálbez, Enhanced thermal energy supply via central solar heating plants with seasonal storage: a multi-objective optimization approach, Appl. Energy 181 (2016) 549–561, https://doi.org/10.1016/j.apenergy.2016.08.037.
[19] E. Oró, M. Codina, J. Salom, Energy model optimization for thermal energy storage system integration in data centres, J Energy Storage 8 (2016) 129–141, https://doi.org/10.1016/j.est.2016.10.006.
[20] A. Prada, A. Gasparella, P. Baggio, On the performance of meta-models in building design optimization, Appl. Energy 225 (2018) 814–826, https://doi.org/10.1016/j.apenergy.2018.04.129.
[21] V. Tulus, M.H. Abokersh, L.F. Cabeza, M. Vallès, L. Jiménez, D. Boer, Economic and environmental potential for solar assisted central heating plants in the EU residential sector: contribution to the 2030 climate and energy EU agenda, Appl. Energy 236 (2019) 318–339, https://doi.org/10.1016/j.apenergy.2018.11.094.
[22] B. Talebi, F. Haghighat, P. Tuohy, P.A. Mirzaei, Optimization of a hybrid community district heating system integrated with thermal energy storage system, J Energy Storage 23 (2019) 128–137, https://doi.org/10.1016/j.est.2019.03.006.
[23] M.H. Abokersh, M. Vallès, L.F. Cabeza, D. Boer, A framework for the optimal integration of solar assisted district heating in different urban sized communities: a robust machine learning approach incorporating global sensitivity analysis, Appl. Energy (2020) 267, https://doi.org/10.1016/j.apenergy.2020.114903.
[24] M.F. Habib, M. Ali, N.A. Sheikh, A.W. Badar, S. Mehmood, Building thermal load management through integration of solar assisted absorption and desiccant air conditioning systems: a model-based simulation-optimization approach, Journal of Building Engineering (2020) 30, https://doi.org/10.1016/j.jobe.2020.101279.
[25] A.X. Naves, L.J. Esteller, A.N. Haddad, D. Boer, Targeting energy efficiency through air conditioning operational modes for residential buildings in tropical climates, assisted by solar energy and thermal energy storage. Case study Brazil, Sustainability (Switzerland) (2021) 13, https://doi.org/10.3390/su132212831.
[26] M. Usman, M. Ali, T. ur Rashid, H.M. Ali, G. Frey, Towards zero energy solar households – a model-based simulation and optimization analysis for a humid subtropical climate, Sustain Energy Technol Assess (2021) 48, https://doi.org/10.1016/j.seta.2021.101574.
[27] G. Wang, G. Feng, X. Li, H. Li, T. Wang, Design optimization integrating energy, economic, and environmental evaluation of a hybrid solar heating system for detached buildings in rural China, Journal of Building Engineering (2023) 73, https://doi.org/10.1016/j.jobe.2023.106692.
[28] T. Yang, W. Liu, Q. Sun, W. Hu, G.J. Kramer, Techno-economic-environmental analysis of seasonal thermal energy storage with solar heating for residential heating in China, Energy 283 (2023) 1–16, https://doi.org/10.1016/J.ENERGY.2023.128389.
[29] B. Li, Z. Liu, Y. Zheng, H. Xie, L. Zhang, Economy and energy flexibility optimization of the photovoltaic heat pump system with thermal energy storage, J Energy Storage (2024) 100, https://doi.org/10.1016/j.est.2024.113526.
[30] Wetter M. GenOpt®-A Generic Optimization Program GenOpt Ö-A Generic Optimization Program. n.d.
[31] A.S. Cruz, L.R. Caldas, V.M. Mendes, J.C. Mendes, L.E.G. Bastos, Multi-objective optimization based on surrogate models for sustainable building design: a systematic literature review, Build. Environ. (2024) 266, https://doi.org/10.1016/j.buildenv.2024.112147.

[32] S.L. Brunton, B.R. Noack, P. Koumoutsakos, Machine learning for fluid mechanics, Annu. Rev. Fluid Mech. 52 (2020) 477–508, https://doi.org/10.1146/annurev-fluid-010719-060214.
[33] W. Huang, J. Xu, Particle swarm optimization, in: Springer Tracts in Civil Engineering, 2023, https://doi.org/10.1007/978-981-99-2213-0_2.
[34] M.D. Shields, J. Zhang, The generalization of Latin hypercube sampling, Reliab. Eng. Syst. Saf. (2016) 148, https://doi.org/10.1016/j.ress.2015.12.002.
[35] W.-L. Loh, On Latin Hypercube Sampling vol. 24, 1996.
[36] M.D. Mckay, R.J. Beckman, W.J. Conover, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code vol. 42, 2000.
[37] L. Liberti, C. Lavor, N. Maculan, A. Mucherino, Euclidean distance geometry and applications, SIAM Rev. 56 (2014) 3–69, https://doi.org/10.1137/120875909.
[38] D.P.P. Mesquita, J.P.P. Gomes, A.H. Souza Junior, J.S. Nobre, Euclidean distance estimation in incomplete datasets, Neurocomputing (2017) 248, https://doi.org/10.1016/j.neucom.2016.12.081.
[39] W.Y. Chiu, G.G. Yen, T.K. Juan, Minimum Manhattan distance approach to multiple criteria decision making in multiobjective optimization problems, IEEE Trans. Evol. Comput. (2016) 20, https://doi.org/10.1109/TEVC.2016.2564158.
[40] M.I. Ribeiro, Gaussian Probability Density Functions: Properties and Error Characterization, 2004.
[41] Clerc Maurice, Particle Swarm Optimization, ISTE Ltd, London, 2006.
[42] S. Amaran, N.V. Sahinidis, B. Sharda, S.J. Bury, Simulation optimization: a review of algorithms and applications, Ann Oper Res 240 (2016) 351–380, https://doi.org/10.1007/s10479-015-2019-x.
[43] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, Global Sensitivity Analysis, The Primer, Chichester, 2008.
[44] M.D. Oisamoje, E.E. Oisamoje, Exploring the economic and environmental benefits of solar energy generation in developing countries: the Nigerian perspective, Journal of Energy Technologies and Policy 3 (2013) 23–32.
[45] S.A. Klein, et al., TRNSYS reference manual, standard component library overview, in: TRNSYS 16, a TRaNsient SYstem Simulation Program 3, 2006.
[46] Hakan İbrahim Tol, TRNSYS From Python. https://github.com/DrTol/TRNSYSfromPython, 2020 (accessed October 9, 2023).
[47] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, Wiley, 2013.
[48] S. Kalogirou, Solar Energy Engineering: Processes and Systems, 2009.
[49] G. Guillen-Gosalbez, J.x.A. Caballero, L.J. Esteller, M. Gadalla, Application of life cycle assessment to the structural optimization of process flowsheets, Computer Aided Chemical Engineering (2007) 24, https://doi.org/10.1016/S1570-7946(07)80218-5.
[50] P. Gluch, H. Baumann, The life cycle costing (LCC) approach: a conceptual discussion of its usefulness for environmental decision-making, Build. Environ. (2004) 39, https://doi.org/10.1016/j.buildenv.2003.10.008.
[51] A. Tarquin, Engineering Economy, El Paso, 2012.
[52] M.S. Peters, K.D. Timmerhaus, R.E. West, Plant Design and Economics for Chemical Engineers, New York, 2003.
[53] M.H. Abokersh, S. Gangwar, M. Spiekman, M. Vallès, L. Jiménez, D. Boer, Sustainability insights on emerging solar district heating technologies to boost the nearly zero energy building concept, Renew. Energy 180 (2021) 893–913, https://doi.org/10.1016/j.renene.2021.08.091.
[54] N.L. Nemerow, F.J. Agardy, P. Sullivan, J.A. Salvato, Environmental Engineering: Environmental Health and Safety for Municipal Infrastructure, Land Use and Planning, and Industry, Sixth edition, 2009, https://doi.org/10.1002/9780470432822.
[55] ISO, 14040: Environmental Management–Life Cycle Assessment—Principles and Framework, International Organization for Standardization, 2006.
[56] S.O. Ryding, ISO 14042: environmental management: life cycle assessment: life cycle impact assessment, Int. J. Life Cycle Assess. (1999) 4, https://doi.org/10.1007/BF02978514.
[57] ISO 14041, Environmental management - life cycle assessment - goal and scope definition and inventory analysis, in: Environmental Management - Life Cycle Assessment - Goal and Scope Definition and Inventory Analysis, 1998, p. 1998.
[58] B. Welsch, L. Göllner-Völker, D.O. Schulte, K. Bär, I. Sass, L. Schebek, Environmental and economic assessment of borehole thermal energy storage in district heating systems, Appl. Energy (2018) 216, https://doi.org/10.1016/j.apenergy.2018.02.011.
[59] Ecoinvent, Ecoinvent database v3. EcoInvent v391, 2023, p. 3.
[60] JRC European Commission, ILCD Handbook: Recommendations for Life Cycle Impact Assessment in the European Context, 2011.
[61] D. Bauer, R. Marx, J. Nußbicker-Lux, F. Ochs, W. Heidemann, H. Müller-Steinhagen, German central solar heating plants with seasonal heat storage, Sol. Energy (2010) 84, https://doi.org/10.1016/j.solener.2009.05.013.
[62] TECNALIA, ACCIONA, DAPPOLONIA, MOSTOSTAL, EFFECTIVE INTEGRATION OF SEASONAL THERMAL ENERGY STORAGE SYSTEMS IN EXISTING BUILDINGS, 2012.
[63] 1.3.5.16. Kolmogorov-Smirnov Goodness-of-Fit Test n.d. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm (accessed October 9, 2023).
[64] F. Sari, Forest fire susceptibility mapping via multi-criteria decision analysis techniques for Mugla, Turkey: a comparative analysis of VIKOR and TOPSIS, For Ecol Manage (2021) 480, https://doi.org/10.1016/j.foreco.2020.118644.
[65] Statistical Institute of Catalonia. Idescat. Population and Housing Census. Dwellings. By type of dwelling. Falset n.d. https://www.idescat.cat/pub/?id=censph&n=30&geo=mun%3A430555&lang=en (accessed October 9, 2023).
[66] Instituto Nacional de Estadística (INE). Censos de Población y Viviendas 2021 n.d. https://www.ine.es/Censo2021/Inicio.do?L=0 (accessed October 9, 2023).
[67] E. de Santiago, J. Arcas-Abella, A. Pagès-Ramon, E. Larrumbide, D. Huerta, Segmentación del parque residencial de viviendas en España en clústeres tipológicos, 2019.
[68] European Union | Eurostat. Energy statistics - cooling and heating degree days n.d. https://ec.europa.eu/eurostat/cache/metadata/en/nrg_chdd_esms.htm (accessed October 9, 2023).
[69] European Commission. JRC Photovoltaic Geographical Information System (PVGIS) - European Commission n.d. https://re.jrc.ec.europa.eu/pvg_tools/es/ (accessed October 9, 2023).
[70] C.T. de la Edificación, Documento Básico he Ahorro de energía, CTE, DB-HE, 2022.
[71] M. Żukowski, Experimental determination of the cold water temperature at the inlet to solar water storage tanks, Thermal Science and Engineering Progress 16 (2020) 1–9, https://doi.org/10.1016/J.TSEP.2019.100466.
[72] A. Pérez-Fargallo, D. Bienvenido-Huertas, S. Contreras-Espinoza, L. Marín-Restrepo, Domestic hot water consumption prediction models suited for dwellings in central-southern parts of Chile, Journal of Building Engineering 49 (2022) 1–13, https://doi.org/10.1016/J.JOBE.2022.104024.
[73] Gobierno de España | Ministerio de Sanidad, Capítulo 3: Sistemas de Agua Caliente Sanitaria [Chapter 3: Domestic Hot Water Systems], Madrid, 2024.
[74] Asociacion de Empresas El'ectricas (ASEME). DATADIS. La plataforma de datos de consumo eléctrico n.d. https://datadis.es/ (accessed October 9, 2023).
[75] European Union | Eurostat. Statistics | Eurostat n.d. https://ec.europa.eu/eurostat/ (accessed October 9, 2023).
[76] Red Electrica. Home | Red Eléctrica n.d. https://www.ree.es/en (accessed October 9, 2023).
[77] Solar Simply con batería virtual | Endesa n.d. https://www.endesa.com/en/catalog/light/tempo/self-consumption-tariff-with-solar-surplus-simply (accessed October 9, 2023).
[78] Trading Economics. EU Natural Gas TTF - Price - Chart - Historical Data - News n.d. https://tradingeconomics.com/commodity/eu-natural-gas (accessed October 9, 2023).
[79] European Commission. Reference and Discount Rates - European Commission n.d. https://competition-policy.ec.europa.eu/state-aid/legislation/reference-discount-rates-and-recovery-interest-rates/reference-and-discount-rates_en (accessed October 9, 2023).
[80] WorldData Info. Inflation Rates in Spain n.d. https://www.worlddata.info/europe/spain/inflation-rates.php (accessed October 9, 2023).
[81] M.H. Abokersh, K. Saikia, L.F. Cabeza, D. Boer, M. Vallès, Flexible heat pump integration to improve sustainable transition toward 4th generation district heating, Energ. Conver. Manage. (2020) 225, https://doi.org/10.1016/j.enconman.2020.113379.
[82] S. Colclough, T. McGrath, Net energy analysis of a solar combi system with Seasonal Thermal Energy Store, Appl. Energy (2015) 147, https://doi.org/10.1016/j.apenergy.2015.02.088.
[83] R. Saez, D. Boer, A.B. Shobo, M. Vallès, Self-consumption potential and surplus compensation policy impact on rooftop photovoltaic systems in Spain, Renew Energy (2024) 229, https://doi.org/10.1016/j.renene.2024.120713.
[84] J. Li, Assessing the accuracy of predictive models for numerical data: not r nor r2, why not? Then what? PloS One (2017) 12, https://doi.org/10.1371/journal.pone.0183250.
[85] Matthias Ehrgott, Multicriteria Optimization, second edition, 2005. Auckland.