Applied Energy 409 (2026) 127469

Contents lists available at ScienceDirect

Applied Energy

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

Improving energy distribution in collective self-consumption via XGBoost-based allocation coefficients prediction

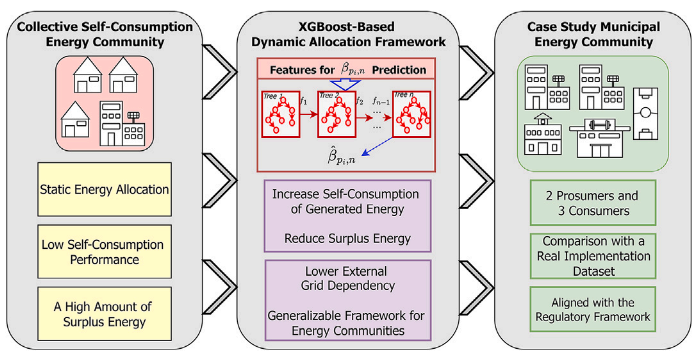
Sebastián Madrigal*, Ramon Gallinad, Jose L. Vicario, Antoni Morell, Ramon Vilanova

Departament de Telecomunicació i Enginyeria de Sistemes, Universitat Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

HIGHLIGHTS

- Developed a data-driven framework for dynamic energy allocation in CSC.
- Applied XGBoost to predict adaptive allocation coefficients in real time.
- Increased self-consumption by 8.4% and reduced surplus energy by 34%.
- Lowered grid dependency by up to 30% across community members.
- Demonstrated scalable, regulation-ready deployment in Spanish ECs.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Energy communities
Collective self-consumption
Energy management
Machine learning for energy systems
XGBoost
Dynamic allocation

ABSTRACT

Energy communities operate under collective self-consumption schemes, where locally generated renewable energy is shared among participating members. In practice, this sharing is commonly governed by static allocation coefficients fixed in advance, which do not capture the time-varying and heterogeneous demand of participants. This mismatch can reduce community self-consumption, increase surplus injections, and raise reliance on the grid. This paper proposes a data-driven framework to dynamically compute allocation coefficients based on predicted individual demand and demonstrates its application in a municipal energy community in Catalonia, Spain. The approach uses an extreme gradient boosting model to forecast hourly consumption profiles and then derive adaptive allocation coefficients that better align shared photovoltaic generation with expected demand. The proposed strategy is evaluated against a static baseline and alternative dynamic schemes using multiple performance indicators, including community self-consumption, surplus energy, and grid dependency. In the case study, the extreme gradient boosting-based allocation increases community self-consumption by 8.4%, reduces surplus energy by 34%, and lowers grid dependency by up to 30% for key members, resulting in a more balanced and efficient distribution of locally generated energy. These results highlight the potential of machine learning-enabled allocation to improve collective self-consumption performance in the existing regulatory framework.

* Corresponding author.
Email addresses: sebastian.madrigal@uab.cat (S. Madrigal), ramon.gallinad@uab.cat (R. Gallinad), jose.vicario@uab.cat (J.L. Vicario), antoni.morell@uab.cat (A. Morell), ramon.vilanova@uab.cat (R. Vilanova).

https://doi.org/10.1016/j.apenergy.2026.127469
Received 16 October 2025; Received in revised form 16 December 2025; Accepted 28 January 2026
Available online 4 February 2026

1. Introduction

Energy Communities (ECs), enable groups of citizens, small enterprises, and public institutions to jointly produce, share, and consume locally generated renewable electricity. By coordinating distributed generation and flexible demand at the local level, ECs support renewable integration, strengthen system resilience, and broaden citizen participation in the energy transition [1–3]. Several operational models have been proposed to implement ECs, including local electricity markets, peer to peer trading, and Collective Self Consumption (CSC) [4,5]. Among these models, the CSC scheme has become one of the most widely adopted approaches in Europe, particularly in Portugal, Spain, Italy, and France, because it provides a practical settlement mechanism through which electricity produced by shared renewable assets, such as rooftop photovoltaic systems, is distributed among members using predefined allocation coefficients.

Despite robust regulatory support across Europe, the operation of CSC-based energy communities faces a persistent technical limitation. Allocation coefficients are typically defined in advance and remain static over extended periods. While this static configuration simplifies administration and billing, it does not reflect the temporal variability of consumption and generation driven by occupancy patterns, weather conditions, and seasonal effects. As a result, mismatches between local generation and demand become recurrent, reducing self-consumption, increasing surplus injections into the grid, and potentially amplifying inequities in the distribution of community benefits [6–8]. To mitigate these inefficiencies, recent studies have proposed dynamic allocation mechanisms, often formulated as optimization-based community energy management problems in which allocation coefficients, or equivalent sharing decisions, are updated to improve efficiency, fairness, or economic performance [9,10]. Although these approaches yield promising results, they commonly depend on detailed modeling assumptions and may entail nontrivial computational effort as community size and operational complexity increase. In parallel, the growing availability of high-resolution metering data has motivated the use of Machine Learning (ML) in community energy management. These data-driven models can capture nonlinear and stochastic relationships in consumption and generation while remaining scalable for real-time or near real-time applications [11,12].

However, a relevant research gap remains. Most ML applications for ECs focus on forecasting, scheduling, or control, whereas the direct prediction of allocation coefficients, which are the operational variables that govern energy sharing in CSC scheme formulation, has received limited attention. Addressing this gap is essential to translate predictive intelligence into actionable and policy compliant sharing decisions.

This study proposes a data-driven allocation framework that infers time varying allocation coefficients from historical smart meter records and operational features. The method relies on Extreme Gradient Boosting (XGBoost) models, selected for their predictive accuracy, computational efficiency, and interpretability when applied to tabular energy data. The framework is evaluated against static and dynamic baselines using community level and participant level performance indicators, and feasibility is demonstrated through a real municipal energy community case study in Catalonia.

The main contributions are summarized as follows:

- Development of a machine learning framework that estimates time varying energy allocation coefficients for CSC-based energy communities, providing a practical and adaptive alternative to static allocation schemes.
- Validation of the framework on a real municipal energy community, demonstrating measurable improvements in self-consumption, surplus reduction, and grid dependency relative to static operation and learning based baselines.
- Implementation of an interpretable and computationally efficient solution based on XGBoost models, supporting scalability and transparency for operational deployment.

The remainder of this paper is organized as follows. Section 2 reviews the literature most relevant to dynamic allocation and data-driven operation in energy communities. Section 3 presents the CSC problem setting and the Spanish regulatory framework for CSC schemes adopted in the case study. Section 4 details the proposed methodology, while Sections 5 and 6 describe the case study and experimental results. Finally, Section 7 summarizes the main findings and outlines future research directions.

2. Related work

This section reviews the contributions most closely related to the problem addressed in this paper. The discussion first examines dynamic allocation and energy sharing mechanisms in Collective Self-Consumption and highlights the regulatory drivers that shape their practical implementation. It then summarizes methodological approaches proposed to overcome the limitations of static coefficients, including optimization based and forecast assisted strategies. Finally, the section reviews machine learning applications in energy communities and positions the present work with respect to the current research gap.

2.1. Dynamic allocation and sharing mechanisms

Collective Self-Consumption has been implemented across Europe through regulatory frameworks that specify proximity requirements, settlement procedures, and the rules by which community generation is allocated among participants through allocation coefficients [13]. These provisions have enabled wide deployment, but they also constrain the design space of sharing schemes. In particular, whether coefficients are required to be fixed or are allowed to vary over time determines the level of operational flexibility, the complexity of implementation, and the types of viable business models. Comparative regulatory analyses indicate that permitting time varying coefficients can improve allocation efficiency and support more advanced community arrangements, albeit with increased requirements for digital infrastructure and operational tooling [14,15].

From a techno-economic perspective, several studies report that static allocation coefficients can yield systematic inefficiencies because they do not adapt to the temporal mismatch between photovoltaic generation and heterogeneous consumption profiles. In the Spanish context, Palou et al. [6] examined the implications of constant distribution coefficients on project profitability and showed that administrative simplicity can come at the cost of underutilizing locally generated energy, which reduces the economic potential of Collective Self-Consumption installations. To mitigate this limitation, the authors proposed alternative sharing schemes that improve profitability, supporting the broader view that allocation flexibility is an important lever for enhancing investment attractiveness in community based photovoltaic projects [16]. Related evidence has been reported in additional Spanish case studies, which indicate that profitability can be constrained under specific regulatory and tariff conditions and that regulatory refinements may be required to strengthen incentives and accelerate adoption [17].

Beyond Spain, demonstration studies have investigated how alternative sharing rules affect operational outcomes. Camblong et al. [8] evaluated multiple allocation modalities in a Collective Self-Consumption pilot in France by comparing static coefficients with dynamic approaches. Their results show that dynamic allocations increase the self-consumption rate and that customized dynamic sharing can yield higher economic savings under different operating scenarios. These findings support the practical value of time varying sharing mechanisms when adequate metering and data acquisition infrastructure is available.

In parallel, optimization-based formulations have been proposed to adjust allocation decisions and improve energy and economic outcomes. The performance in [7] studied Collective Self-Consumption in residential buildings under Spanish regulation and assessed the influence of system sizing, demand shifting through smart appliances, and energy sharing strategies. Their results indicate that the marginal value of dynamic sharing can depend on the presence of other flexibility options, such as storage and demand response, highlighting that the benefits of allocation flexibility are configuration dependent. More recent work reports stronger improvements when dynamic allocation is coupled with predictive information and optimization-based decision-making. The authors in [10] compared coefficient strategies aligned with Spanish regulation against dynamic strategies driven by day-ahead consumption forecasts and reported improvements in self-consumption and financial indicators, particularly when battery energy storage is available. In addition, bi-level formulations have been proposed to represent the interaction between a community level decision maker and participant level objectives [9]. In [18,19], the authors developed bi-level optimization frameworks for Renewable Energy Communities (RECs) that maximize shared benefits while accounting for individual cost minimization and showed that such formulations can remain computationally viable for online operation under realistic Italian energy community sizes.

Overall, the literature indicates broad agreement that static allocation coefficients can limit the efficiency and economic potential of CSC, whereas dynamic allocation offers a pathway to improved self-consumption and participant benefits. However, many existing approaches rely on explicit optimization models that require detailed system representations, careful specification of objectives and constraints, and increasing computational effort as the number of members and operational degrees of freedom grows. These characteristics motivate only data-driven alternatives that provide allocation flexibility with reduced modeling burden while remaining compatible with policy requirements.

2.2. Machine learning applications in energy communities

The deployment of smart meters, distributed energy resources, and connected devices has accelerated the adoption of data-driven methodologies for community level energy management. In this context, ML has been used to improve forecasting accuracy, enable adaptive decision-making, and support near real-time operation through predictive analytics and automation [20]. Existing surveys report that supervised learning is widely applied to regression and classification tasks, whereas reinforcement learning is increasingly adopted for online scheduling and control when sequential decisions under uncertainty are central [12]. A first research stream applies machine learning to support the design and planning of RECs. In [21] a data-driven framework is integrated with life cycle cost and life cycle assessment metrics to guide the sizing of photovoltaic generation, wind turbines, and battery energy storage, and combines it with multi objective optimization and multi criteria decision analysis to support stakeholder decisions. Complementarily, Piras et al. [22] introduced an open tool that leverages artificial intelligence and machine learning techniques to automate renewable energy community configuration and generate optimized sizing scenarios, with the goal of lowering technical barriers and accelerating deployment.

A second stream focuses on the operational layer, where machine learning models provide forecasts or decision support for balancing, trading, and asset scheduling. In [23], XGBoost models are used to forecast electricity consumption, and a decision algorithm exploits forecasts of consumption and solar production together with storage states and market prices to guide interactions with the public grid, reporting reductions in community electricity expenses and dependence on centralized supply under multiple tariff and generation scenarios. In peer to peer settings, Boumaiza and Maher [24] developed a machine learning enhanced home energy management framework that combines scheduling, bid optimization, and real-time adjustment, showing improved economic outcomes and reduced grid dependency in simulated prosumer networks under uncertainty.

A third direction addresses data-driven control for real time community operation, with emphasis on reinforcement learning. In [25], a reinforcement learning controller is proposed for battery energy storage systems in renewable energy communities operating under incentive schemes for virtual Self-Consumption and demonstrates performance improvements relative to rule based controllers while relying only on information available at the current time step. Extending this line of work, Palma et al. [26] reported a European scale study showing that reinforcement learning agents can generalize across multiple energy community configurations and market contexts, achieving robust performance comparable to benchmark controllers in simulation. At a broader system level, Gonçalves et al. [27] introduced an ML as a service framework in which customized reinforcement learning agents are deployed per community microgrid, supported by forecasting and scalable training and inference pipelines, and reported cost reductions relative to heuristic baselines while enabling extensions such as peer to peer trading and internal energy pooling.

Taken together, these contributions establish ML as an increasingly practical enabler of intelligent community operation across planning, forecasting assisted management, and real-time control. However, most existing applications emphasize forecasting and decision support for dispatch, trading, and storage control, rather than learning the energy sharing variables that directly determine how CSC generation is allocated among participants. This motivates targeted data-driven approaches that translate learned patterns into operational allocation decisions under CSC operational constraints.

2.3. Positioning and research gap

The energy community literature emphasizes that operational performance depends on the interaction between technical design, policy constraints, and business models through which collective benefits are realized. Comprehensive reviews highlight that modeling choices and optimization objectives are often addressed separately, leaving research opportunities at the intersection of technical operation, regulatory mechanisms, and value creation [28]. In parallel, machine learning surveys show that data-driven contributions in local energy communities primarily address forecasting, energy management systems, storage optimization, and transaction support, with supervised learning used for prediction tasks and reinforcement learning for control oriented applications [11].

Building on these insights, a specific limitation emerges for CSC-based energy communities. The allocation layer has received comparatively limited attention from a data-driven perspective. Although energy sharing has been studied using static coefficients and, more recently, optimization based coordination, ML is most often used to forecast consumption or generation rather than to learn the allocation variables that directly govern energy distribution among participants. This work addresses this gap by introducing a policy aligned predictive framework that estimates dynamic allocation coefficients directly from operational data, enabling adaptive energy sharing with computational efficiency and interpretability suitable for community scale deployment.

3. Foundations of energy communities

The ongoing transformation of energy systems toward decentralization and sustainability has increased attention on the role of energy communities. These initiatives enable citizens, small enterprises, and public institutions to jointly produce, share, store, and consume energy, primarily from renewable sources such as solar photovoltaics (PV). The European Union formally recognizes ECs within the Clean Energy Package as a means to democratize energy access and foster community-level resilience. Importantly, the primary purpose of these communities is not financial profit, but the delivery of social, environmental, and local economic benefits.

3.1. Collective self-consumption: the Spain formulation

In Spain, these concepts have materialized through the CSC distribution scheme. Institutionalized through Royal Decree Law and Royal Decree provisions [29,30], the CSC scheme enables multiple consumers to share locally generated renewable energy (mainly solar PV), within a defined geographical area. Proximity rules require that generation and consumption points lie within 500 m, although this distance can extend to 5000 m under specific urban conditions. This legal structure reinforces the local character of energy production and usage, while supporting grid stability and reducing transmission losses.

The Collective Self-Consumption scheme has become Spain's leading regulatory instrument for enabling energy communities. It allows participants who cannot individually deploy generation systems, such as apartment dwellers or small organizations, to benefit from shared renewable energy. A CSC arrangement typically requires a formal agreement among participants and, in many cases, a coordinating entity responsible for administrative procedures and interactions with the grid. Spanish law defines four operational modalities of CSC, which differ in how surplus energy is treated:

- **CSC without surplus:** all generated energy is consumed instantaneously using an anti-spill device, and grid injection is not allowed.
- **CSC without surplus with simplified compensation:** surplus is not physically exported, but it can be virtually redistributed among users for internal balancing.
- **CSC with surplus and simplified compensation:** excess energy is injected into the grid and compensated on the electricity bill, limited to the user grid consumption.
- **CSC with surplus without simplified compensation:** surplus is sold on the wholesale market, which requires market registration and is generally suited to larger cooperatives.

For CSC schemes to operate effectively, participants take on well defined roles. Prosumers are members who both generate and consume electricity, often co-financing the PV installation. Consumers only receive shared energy without participating in generation ownership. In addition, third-party coordinators, either internal participants or external service providers, manage allocation processes, ensure regulatory compliance, and serve as the interface with the Distribution System Operator (DSO). Although Spanish regulations allow the integration of energy storage systems, storage is not typically coordinated at the community level. Instead, storage remains under individual control, which can limit the potential to optimize energy flows across the community as a whole.

3.2. Mathematical formulation of energy distribution in CSC

A core regulatory feature of the CSC scheme is the employment of allocation coefficients, denoted β and studied in this work, to determine the share of generated energy received by each user. These coefficients are legally required and must be submitted to the DSO. This mechanism makes CSC operation quantifiable, administrable, and verifiable for distribution companies and regulatory authorities.

Let $\mathrm{TEG}(t)$ denote the total energy generated for CSC allocation at time step t. The individualized net energy assigned to member n at time t is then:

$$\mathrm{INE}_n(t) = \beta_n(t)\mathrm{TEG}(t), \tag{1}$$

where $\mathrm{INE}_n(t)$ is the Individualized Net Energy assigned to member n, $\beta_n(t)$ is the corresponding allocation coefficient, and $\mathrm{TEG}(t)$ is the total CSC-eligible energy generated at time t. This formulation ensures that each participant receives a proportionate share of energy according to pre-agreed criteria, including consumption needs, financial investment, or other equity principles. To distribute all generated energy without loss or duplication, the coefficients must satisfy the conservation constraint:

$$\sum_{n=1}^{N} \beta_n(t) = 1, \tag{2}$$

where N is the number of members in the CSC community.

3.3. Understanding allocation coefficients

Spanish regulation allows two types of allocation coefficients in CSC schemes:

- **Static coefficients:** Coefficients are fixed over time and are commonly adopted due to administrative simplicity. Allocation is typically based on predefined shares, such as investment participation or historical consumption.
- **Dynamic coefficients:** Coefficients may vary at hourly resolution to reflect time-varying usage patterns. Their implementation requires advanced infrastructure, including smart metering, reliable data exchange, and dedicated software for coefficient estimation and submission.

As outlined in Order [31], dynamic coefficients are legally accepted when submitted before each billing cycle. Despite improved alignment between PV generation and demand, adoption remains limited due to technical, administrative, and financial barriers.

Key challenges include the lack of standardized digital tools, limited technical expertise, and restricted community engagement [32]. Nevertheless, dynamic allocation remains a promising mechanism to improve both efficiency and equity in CSC. As digital infrastructure and operational platforms mature, dynamic schemes are expected to become more feasible, enabling energy communities to more effectively leverage local renewable energy resources.

Nomenclature: The acronyms and symbols used throughout the work are provided in Appendix A.

4. Methodology

This section presents a data-driven methodology to improve energy allocation in Collective Self-Consumption (CSC) communities while remaining compliant with operational constraints. The proposed approach learns time-varying allocation coefficients using XGBoost regressors. These models are trained to approximate ideal allocation targets that reduce the mismatch between allocated energy and member demand and, consequently, limit surplus injections. Training relies on historical consumption and photovoltaic generation data augmented with calendar descriptors and lagged operational variables, which enables hour ahead coefficient estimation with low computational burden. The remainder of this section describes the pipeline, the learning and validation procedure, and the post-processing steps used to enforce feasibility and energy conservation before applying the learned policy in energy community operation.

4.1. Adoption of XGBoost

Dynamic allocation in the CSC scheme requires high predictive accuracy on tabular features, low inference latency, and sufficient transparency to support deployment and auditing. XGBoost regressors meet these requirements and have shown strong performance in energy applications where nonlinear effects and seasonal structures are relevant [33,34]. In this work XGBoost is adopted as a practical compromise between accuracy and computational cost, while retaining interpretability through feature-importance analysis and tree-based explanations.

XGBoost models are employed as an allocation learners rather than as a pure forecasting tool. Given calendar descriptors and lagged consumption and generation variables, the model maps the observed community state to time-varying allocation coefficients that approximate an ideal proportional allocation target. This differs from common community-energy pipelines in which ML is used only to forecast demand or generation as an input to an optimization layer. By learning allocation variables directly, the proposed approach reduces modeling burden and supports scalable hourly coefficient updates. Feasibility is enforced through post-processing normalization that guarantees energy conservation.

4.2. Ideal proportional heuristic framework

Consider a CSC-based energy community composed of two disjoint sets of participants. The set of consumers is $\mathcal{C} = \{c_1, \dots, c_K\}$ and the set of prosumers is $\mathcal{P} = \{p_1, \dots, p_I\}$, with $\mathcal{C} \cap \mathcal{P} = \emptyset$. Consumers exclusively demand electricity and follow consumption profiles $C_k(t)$. Each prosumer p_i generates photovoltaic energy $G_i(t)$ and consumes electricity $C_i(t)$. The set of all members is $\mathcal{N} = \mathcal{C} \cup \mathcal{P}$ with cardinality $N = |\mathcal{N}| = I + K$. For any member $n \in \mathcal{N}$, let $C_n(t)$ denote electricity consumption at time t, and collect all member consumptions as $\mathbf{C}(t) = [C_1(t), \dots, C_N(t)]$. Total community PV generation is $G_{\text{tot}}(t) = \sum_{i=1}^{I} G_i(t)$. This prosumer and consumer distinction is consistent with the Spanish CSC framework, where participants may simultaneously contribute generation and consumption while remaining analytically separable in the energy community model.

To construct supervised-learning targets, an ideal allocation rule is defined that distributes PV energy proportionally to contemporaneous demand, this rule was previously applied in [35] as a proposal to improve energy allocation by reducing surplus energy. This heuristic captures the intended behavior of dynamic allocation coefficients by prioritizing members with higher instantaneous demand, and it provides a consistent target $\beta^*_{p_i,n}(t)$ for training. The ideal coefficients depend on realized consumption at time t, which is only known ex-post after the metering interval closes. Therefore, this rule cannot be applied online unless future consumption is forecasted. In Section 4.3, a predictive approximation of these coefficients is learned using only information available at decision time, namely lagged metering variables and calendar descriptors, so that allocation coefficients can be produced operationally without access to $C_n(t)$.

The Algorithm 1 computes the target ideal proportional allocation coefficients. The term $d_n(t) = \min\big(C_n(t), G_{\text{tot}}(t)\big)$ denotes the effective demand of member n, and $D(t) = \sum_{m=1}^{N} d_m(t)$ is the corresponding aggregate effective demand. By construction, the ideal coefficients satisfy the CSC conservation constraint for each prosumer:

$$\sum_{n=1}^{N} \beta^*_{p_i,n}(t) = 1 \qquad \forall\, i \in \mathcal{P} \tag{3}$$

Moreover, they increase the allocated share for members with higher demand at time t. Community-level surplus can still occur when total PV generation exceeds aggregate community demand. In those hours, surplus is unavoidable because the energy distributed cannot be fully absorbed by members. The heuristic cannot eliminate surplus in generation-dominant periods, but it aims to reduce it whenever feasible by allocating energy in proportion to demand while enforcing a CSC-compliant settlement rule.

Algorithm 1 Ideal proportional allocation employed as training target.

Require: $\{C_n(t)\}_{n=1}^{N}$, $\{G_i(t)\}_{i=1}^{I}$
Ensure: $\{\beta^*_{p_i,n}(t)\}$ for all i, n

$G_{\text{tot}}(t) \leftarrow \sum_{i=1}^{I} G_i(t)$
if $G_{\text{tot}}(t) = 0$ **then**
    for $i = 1, \dots, I$ **do**
        for $n = 1, \dots, N$ **do**
            $\beta^*_{p_i,n}(t) \leftarrow \frac{1}{N}$
        end for
    end for
else
    for $n = 1, \dots, N$ **do**
        $d_n(t) \leftarrow \min\big(C_n(t),\, G_{\text{tot}}(t)\big)$
    end for
    $D(t) \leftarrow \sum_{m=1}^{N} d_m(t)$
    for $i = 1, \dots, I$ **do**
        for $n = 1, \dots, N$ **do**
            $\beta^*_{p_i,n}(t) \leftarrow \frac{d_n(t)}{D(t)}$
        end for
    end for
end if

4.3. Learning to predict allocation coefficients

The objective of this stage is to learn a mapping from information available at decision time to time-varying CSC allocation coefficients that approximate the ideal proportional targets defined in Section 4.2. The Fig. 1 summarizes the workflow. Historical metering records are first transformed into calendar descriptors and lagged operational features. The required inputs are the time series of electricity consumption for each member, $C_n(t)$, and PV generation for each prosumer, $G_i(t)$, over the training horizon. Ideal coefficients $\beta^*_{p_i,n}(t)$ are then computed offline from realized consumption and PV generation using Algorithm 1. These targets are paired with the corresponding feature vectors to train the XGBoost regressors. During operation, the trained models take the

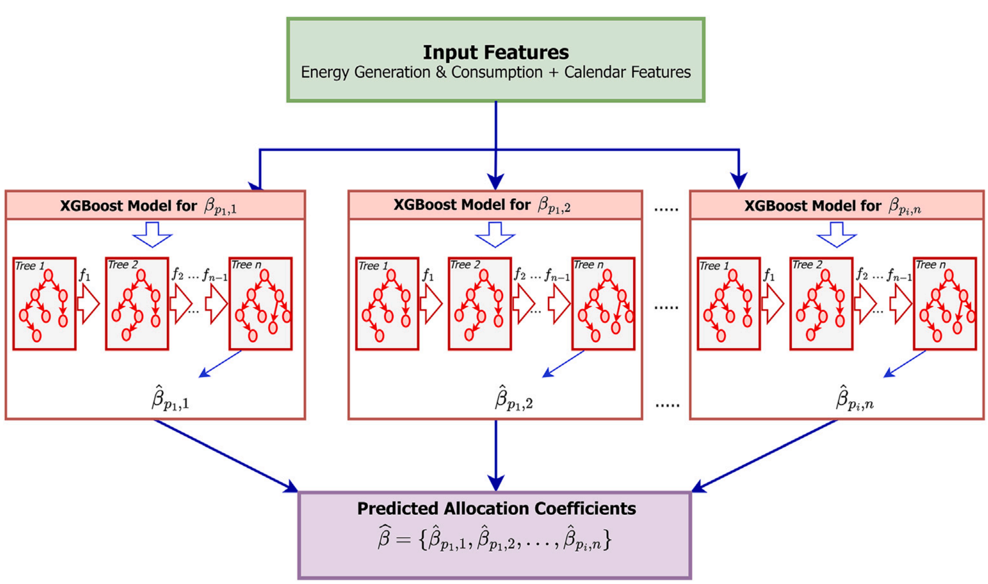


Fig. 1. Data flow for training and deployment of time-varying allocation coefficients. Calendar descriptors and lagged metering variables are used as inputs to an ensemble of XGBoost regressors that estimate the coefficients set.

current feature vector and output coefficient estimates that are post-processed to enforce feasibility and energy conservation. The input variables are designed to be compatible with standard CSC metering practices. Consumption measurements $C_n(t)$ are assumed to be available for all members, and PV generation measurements $G_i(t)$ are available for prosumers. Calendar descriptors are constructed from the timestamp to represent periodic structures in community operation, including diurnal and weekly patterns, seasonal effects, and the influence of weekends and public holidays depending on the country. Lagged variables are included for both consumption and PV generation to capture short-term persistence without relying on exogenous measurements. In the implementation, lags of 1, 24, and 168 hours are employed. This requires an initial history window, therefore the first 168 hours are used to initialize the lagged feature set.

Supervision is obtained by generating ideal proportional targets offline. For each time t, Algorithm 1 computes $\beta^*_{p_i,n}(t)$ from realized metering data. These targets are used only to generate training labels and to provide an offline performance reference. Their construction does not imply perfect foresight at deployment. Instead, they define a consistent supervision signal for learning a mapping based on lagged measurements and calendar information, which are available at decision time.

The learning task is implemented as a bank of regressors, one per coefficient $\beta_{p_i,n}(t)$. Each model receives the same feature categories, combining calendar descriptors with lagged consumption and lagged PV generation. The predicted coefficient for each prosumer member pair is expressed as:

$$\hat{\beta}_{p_i,n}(t) = f_\theta\left(x^{\mathrm{cal}}(t),\, x_n^{\mathrm{lags}}(t),\, x_{p_i}^{\mathrm{lags}}(t)\right) \tag{4}$$

where $x^{\mathrm{cal}}(t)$ denotes the calendar feature vector, and $x_n^{\mathrm{lags}}(t)$ and $x_{p_i}^{\mathrm{lags}}(t)$ collect lagged consumption and generation variables, respectively. Training and validation follow time-consistent splitting to avoid information leakage across time. The resulting models are retained for hour-ahead inference.

Since independent regression does not guarantee the exact satisfaction of the conservation constraint, feasibility is enforced through post-processing. For each prosumer p_i and time t, raw predictions are normalized as:

$$\tilde{\beta}_{p_i,n}(t) = \frac{\hat{\beta}_{p_i,n}(t)}{\sum_{n=1}^{N} \hat{\beta}_{p_i,n}(t)} \tag{5}$$

which ensures:

$$\sum_{n=1}^{N} \tilde{\beta}_{p_i,n}(t) = 1 \qquad \forall\, i \in \mathcal{P} \tag{6}$$

The normalized coefficients $\tilde{\beta}_{p_i,n}(t)$ define the operational allocation policy employed in the subsequent evaluation and simulation under the CSC scheme.

4.4. Framework scalability overview

The proposed methodology is intended to support community-scale deployment with hourly coefficient updates. Scalability is primarily driven by the number of allocation coefficients that must be estimated, since the learning task is implemented as a bank of regressors followed by a lightweight normalization step.

Let N denote the total number of members, with I prosumers and K consumers such that $N = I + K$. Under the prosumer-centered definition adopted in this work, each prosumer allocates its generation across all members, yielding a total number of coefficients:

$$B = I \cdot N = I \cdot (I + K) \tag{7}$$

This expression shows that complexity increases linearly with the number of consumers and quadratically with the number of prosumers. Adding a consumer increases B by I. Adding a prosumer increases B by $N + I$ because the new prosumer introduces coefficients toward all existing members and also appears as a recipient in the coefficient sets of all prosumers. The Fig. 2 illustrates this scaling behavior for representative membership changes. The same pipeline supports periodic retraining or incremental updating as new metering data become available, which allows the learned the allocation policy to adapt over time without modifying the underlying CSC-compliant structure.

4.5. Energy community performance metrics

Performance is evaluated through technical indicators that quantify community PV utilization and the alignment between allocated energy and member demand. Let $C_n(t)$ denote the electricity consumption of member n at time t, and let $D_n(t)$ denote the energy allocated to member n at time t through the CSC allocation coefficients. All metrics are computed over a horizon of T hourly samples.

Energy surplus. Surplus energy at time t is defined as the portion of allocated energy that cannot be absorbed by aggregate demand and is therefore exported:

$$\mathrm{ES}(t) = \max\left[\sum_{n=1}^{N} D_n(t) - \sum_{n=1}^{N} C_n(t),\, 0\right] \tag{8}$$

Lower $\mathrm{ES}(t)$ indicates improved alignment between distribution and demand and reduced injections into the grid.

Energy balance. For member-level inspection, the instantaneous balance is defined as:

$$\mathrm{EB}_n(t) = C_n(t) - D_n(t) \tag{9}$$

which is positive under under-allocation and negative under over-allocation. This index is used to visualize allocation mismatches over time.

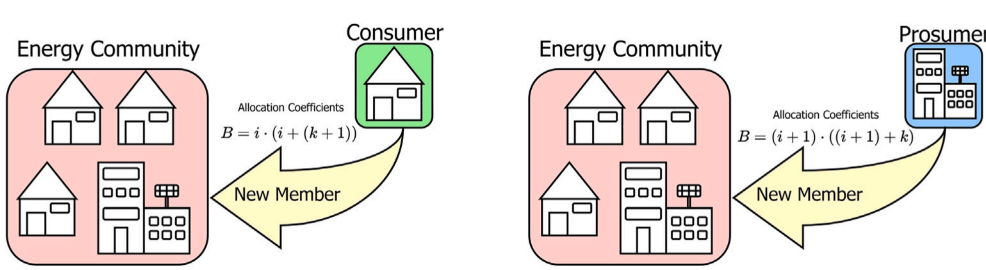


Fig. 2. Effect of adding a new community member on the number of CSC allocation coefficients under the prosumer-centered definition. Adding a consumer introduces one additional coefficient per existing prosumer, whereas adding a prosumer increases the coefficient set more strongly because the new member acts as both an energy provider and an energy receiver.

Surplus frequency index. The frequency of surplus events is:

$$\mathrm{SFI} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}[\mathrm{ES}(t) > 0] \tag{10}$$

Average surplus per event. The average magnitude of surplus conditional on a surplus event is:

$$\mathrm{ASE} = \frac{\sum_{t=1}^{T} \mathrm{ES}(t)\,\mathbf{1}[\mathrm{ES}(t) > 0]}{\sum_{t=1}^{T} \mathbf{1}[\mathrm{ES}(t) > 0]} \tag{11}$$

Together, SFI and ASE distinguish frequent small exports from infrequent large exports.

Receiver utilization index. Receiver utilization for member n is defined as:

$$\mathrm{RUI}_n = \frac{\sum_{t=1}^{T} \min\left[D_n(t),\, C_n(t)\right]}{\sum_{t=1}^{T} D_n(t)} \tag{12}$$

Higher RUI_n indicates that allocated energy is largely consumed locally rather than exported.

Grid dependency ratio. Grid dependency for member n is defined as:

$$\mathrm{GDR}_n = \frac{\sum_{t=1}^{T} \max\left[C_n(t) - D_n(t),\, 0\right]}{\sum_{t=1}^{T} C_n(t)} \tag{13}$$

Lower GDR_n indicates higher coverage of demand through community PV sharing.

Energy balance accuracy. Allocation accuracy for member n is measured as the mean absolute deviation of the balance:

$$\mathrm{EBA}_n = \frac{1}{T}\sum_{t=1}^{T} \left|\mathrm{EB}_n(t)\right| \tag{14}$$

Smaller EBA_n indicates tighter matching between allocated energy and realized demand.

Energy-weighted mean absolute error. To prioritize coefficient accuracy during high PV hours, an energy-weighted mean absolute error is defined as:

$$\mathrm{WMAE} = \frac{\sum_{t=1}^{T} \omega_t \left|\beta^*(t) - \hat{\beta}(t)\right|}{\sum_{t=1}^{T} \omega_t} \tag{15}$$

where ω_t denotes PV availability at time t, and $\beta^*(t)$ and $\hat{\beta}(t)$ denote the ideal and predicted allocation coefficients under evaluation. This metric emphasizes errors during hours in which allocation decisions have the largest impact on surplus and self-consumption.

4.6. Surplus compensation management

Under the Spanish CSC scheme, surplus energy that cannot be absorbed by members at hour t is injected into the public grid as surplus and valued according to the compensation mechanism specified by the applicable CSC modality and the retailer contract. In the assessment, this mechanism is represented through an exogenous time series of compensation prices, denoted $\pi_{\mathrm{comp}}(t)$, which determines the monetary value of exported energy over the evaluation horizon.

Compensation conditions differ from retail tariffs and may be low and time-varying. Consequently, surplus export can contribute limited revenues relative to the value of locally self-consumed energy. This motivates allocation strategies that reduce surplus whenever feasible within CSC operational constraints.

The Fig. 3 reports the hourly evolution of $\pi_{\mathrm{comp}}(t)$ for 2023, obtained from publicly available market datasets [36]. The series exhibits substantial temporal variability and includes intervals in which exporting surplus provides limited economic value. In the results section, this time dependence is used to quantify the sensitivity of annual revenues to the timing of surplus, not only to its magnitude.

The following Fig. 4 complements the time series by summarizing the empirical distribution of $\pi_{\mathrm{comp}}(t)$ across the year. This representation provides a compact view of typical compensation levels and dispersion and supports the interpretation of economic outcomes by relating changes in surplus energy to the prevailing compensation regime over the evaluation horizon. Together, these figures motivate the emphasis on reducing surplus injections as a primary operational objective, since exported energy is valued under a volatile and contract-dependent compensation mechanism.

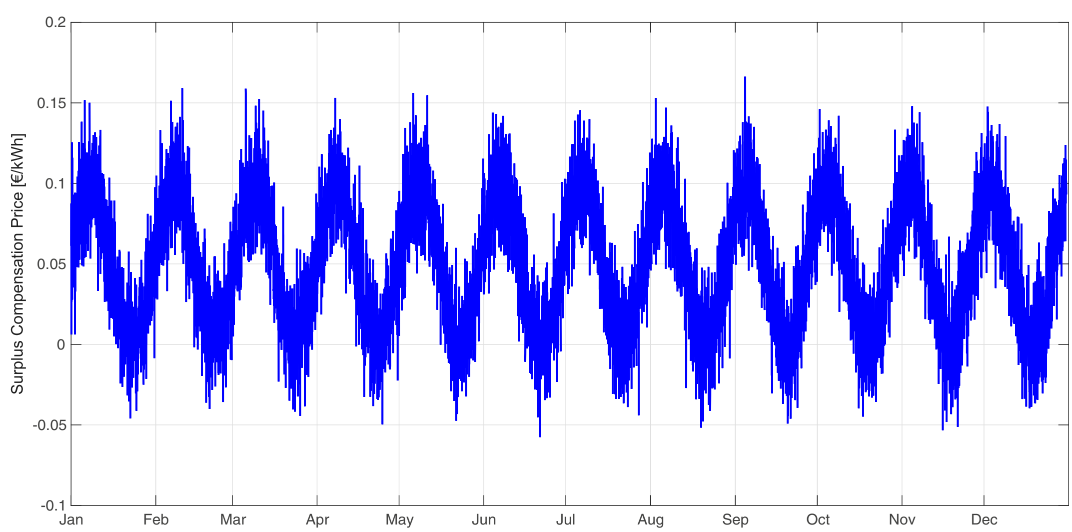


Fig. 3. Hourly surplus compensation price in Spain during 2023, $\pi_{\mathrm{comp}}(t)$, used to value exported surplus in the economic assessment [36].

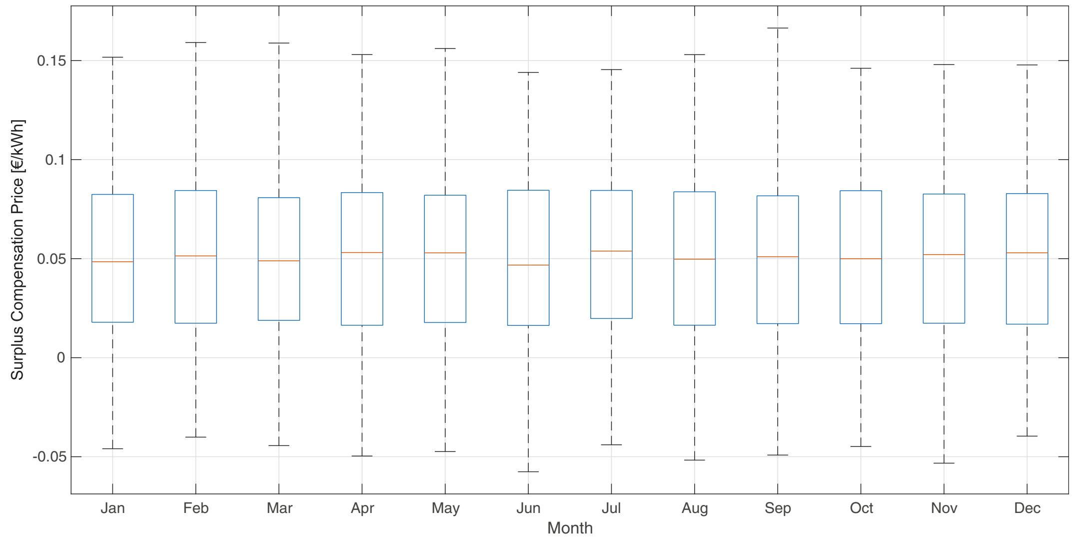


Fig. 4. Monthly distribution of hourly surplus compensation prices in Spain during 2023 [36]. Box plots show typical levels and variability by month.

5. Case study: municipal energy community in catalonia

The case study considers a real Municipal Energy Community (MEC) located in the province of Barcelona, Spain. Due to confidentiality requirements, municipality-specific identifiers and sensitive operational details are not disclosed. The MEC operates under the Spanish Collective Self-Consumption regulatory framework and represents an increasingly common arrangement in which local governments coordinate the operation of shared PV generation across municipal buildings, with the possible inclusion of citizen participation. This type of initiative aligns with European objectives related to decentralization, energy democracy, and climate neutrality, while also addressing local priorities such as public-sector sustainability and the mitigation of energy poverty.

The community includes two schools acting as prosumers, each equipped with rooftop PV. Generated energy is first used on site and the CSC-eligible share is then allocated to other municipal buildings, namely a sports center, a sports area, and a civic center, which act solely as consumers. In addition, a portion of the generation is assigned to a collective participant denoted as *citizenship*, which represents local households that contributed financially to the installations and are entitled to a fixed share of the PV output. The MEC operates under the CSC with surplus and simplified compensation modality, whereby unused energy is exported to the grid and financially compensated. From an operational standpoint, the case study benefits from centralized decision-making through the city hall. Since the municipality owns the participating buildings and coordinates administrative procedures and billing, changes in allocation strategy can be implemented with reduced contractual friction, which provides an appropriate setting for assessing dynamic allocation methods based on predictive models.

5.1. Nominal case technical and data description

A schematic representation of the community is provided in Fig. 5, which illustrates the two generation sites, School E and School P, and the corresponding allocation of PV energy toward the institutional receivers and the citizenship component. The configuration reported in this section corresponds to the current implementation of the MEC, including the self-consumption fraction and the static allocation coefficients used in practice. The technical infrastructure includes two rooftop PV systems with installed capacities of 140 kWp and 86 kWp. The Table 1 summarizes these parameters and the self-consumption fraction α, which is set to 70% for both generators in accordance with the nominal operational configuration.

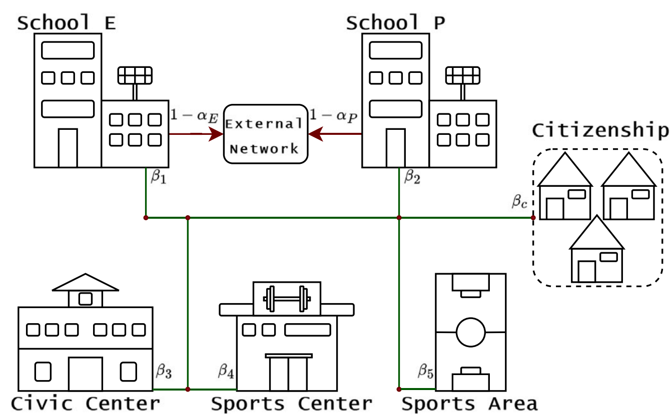


Fig. 5. Schematic of the studied MEC under the Spanish CSC scheme, showing prosumer PV generation, member allocation coefficients β, and the self-consumption fraction α.

Table 1
Photovoltaic generator parameters used in the nominal scenario for each prosumer building, including installed capacity and the assumed self-consumption fraction α.

Generator	School E	School P
Installed Power (kWp)	140	86
Self-Consumption Fraction (%)	70	70

The analysis relies on hourly measurements spanning the full calendar year 2023, including electricity consumption for each of the five institutional participants and PV generation for each of the two prosumers. Although the citizenship component participates in the allocation process, individual residential consumption profiles are not available. To ensure regulatory consistency and preserve the contractual structure of the MEC, the citizenship share is treated as a fixed withdrawal from the CSC-eligible PV energy before any prediction or reallocation among institutional members is performed.

Under this configuration, only 25.4% of the CSC-eligible energy is available for optimization among the five institutional receivers, which corresponds to 53,226 kWh per year. The remaining 74.6% of the self-consumed share is reserved for the citizenship component based on the pre-assigned participation packages. In Table 2 the nominal allocation

Table 2
Nominal CSC allocation coefficients used for the case study, reporting the percentage of PV energy from each prosumer assigned to each receiver and the assumed self-consumption fraction α.

Receiver	From School E (%)	From School P (%)	Self-Consumption Fraction (%)
School E	9.00	0.00	70
School P	0.00	17.80	70
Sports Area	14.45	0.00	70
Sports Center	6.55	0.00	70
Civic Center	0.00	0.22	70
Citizenship	70.00	81.98	70

coefficients are reported, reflecting this fixed split. The hourly resolution of the dataset enables a fine-grained assessment of both the baseline and the proposed dynamic allocation strategies. The participating buildings exhibit heterogeneous consumption dynamics driven by operating schedules, seasonal activity, and usage intensity, which is essential for evaluating the responsiveness of time-varying coefficients relative to static allocation.

The Fig. 6 reports the annual consumption totals of the institutional members together with the total PV generation under the CSC configuration. The Sports Area is the largest consumer with more than 90,000 kWh per year, while the Civic Center exhibits the lowest demand, slightly above 3600 kWh. The aggregated annual electricity demand of the institutional participants is close to 300,000 kWh, which far exceeds the portion of PV generation that is eligible for distribution to these receivers under the fixed self-consumption fraction and the citizenship withdrawal. The baseline scenario corresponds to the current operation of the MEC, in which each building receives a fixed percentage of energy from each generator regardless of contemporaneous demand. For example, the Sports Area receives 14.45% of School E generation and none from School P, while the Civic Center receives 0.22% from School P. This allocation is legally compliant but does not adapt to demand variability or occupancy schedules. During low-demand periods, including weekends and holiday intervals, a portion of the allocated energy may remain unused and is therefore exported to the grid, which can reduce both operational efficiency and economic benefit.

In the proposed dynamic scenario, an XGBoost model is used to infer hour-ahead allocation coefficients directly from calendar variables and lagged metering features. Predicted coefficient vectors are post-processed to satisfy feasibility constraints and to preserve the fixed citizenship share. The simulations assume no battery storage, no changes to the low-voltage distribution network, and PV generation equal to the recorded measurements. Any energy that cannot be absorbed by institutional demand after CSC allocation is exported to the grid under the simplified compensation modality and valued using the corresponding day-ahead compensation price. This operational representation reflects the logic commonly applied in Spanish CSC deployments and supports a realistic comparison between static and dynamic allocation.

5.2. Economical considerations of the PV installation

The PV infrastructure supporting the MEC represents a substantial capital investment by the municipality. The combined installed capacity of 140 kWp at School E and 86 kWp at School P yields a total of 226 kWp. Based on indicative market costs for grid-connected PV installations in Spain in 2025, reported in the range from 1200€ to 1600€ per kWp including value-added tax, engineering, and permitting [37], the total investment is estimated to be between 271,200€ and 361,600€. Since the installation was completed prior to 2022, the realized costs may differ from this range due to temporal variations in equipment and installation prices.

Financial incentives can substantially reduce the net investment. National and European programs often provide investment support on the order of 600€ per kWp, while some regional schemes for public buildings can reach higher levels. Using a conservative value of 600€ per kWp, the total subsidy would be approximately 135,600€, which yields an estimated net investment between 135,600€ and 226,000€. In this context, the self-consumption fraction of 70% is an important determinant of economic viability. Under Spanish CSC settlement, exported surplus is compensated, typically at a lower value than the avoided retail cost of imported electricity. Therefore, increasing internal consumption provides direct bill savings, improves the effective value of PV generation, and contributes to shorter payback periods.

The residual capital requirement, likely in the range from 150,000€ to 250,000€, can be financed through a combination of municipal funding instruments, low-interest green loans, and citizen contributions. In the studied MEC, citizen participation is represented through a crowdfunding mechanism in which households co-finance the PV systems and receive entitlements to a fixed share of the PV output based on participation packages expressed in installed capacity. This financing structure supports economic sustainability while strengthening citizen engagement and alignment with municipal climate and energy objectives.

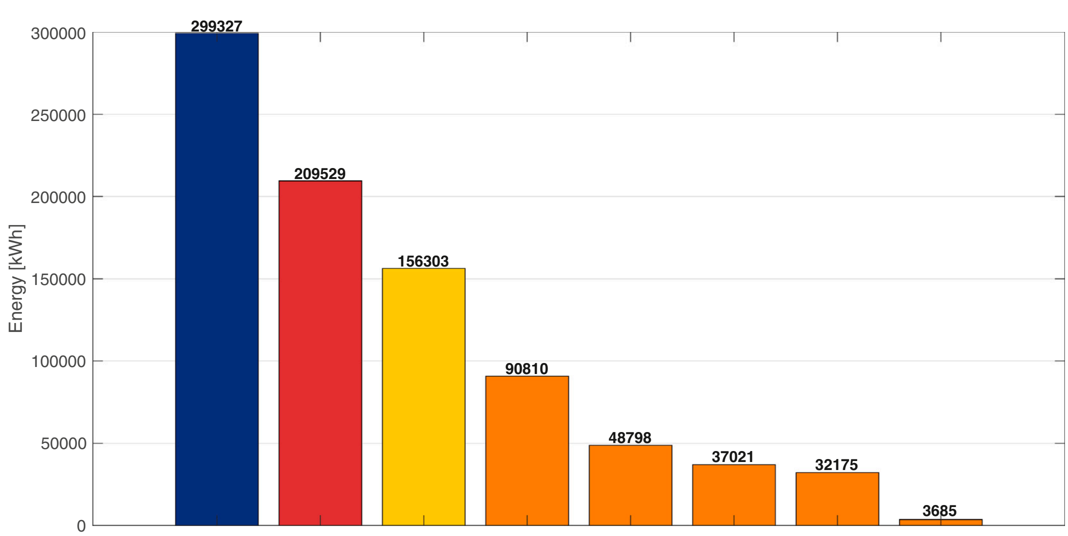


Fig. 6. Annual energy totals for the MEC, comparing aggregated electricity demand by member and total PV generation under the CSC configuration, highlighting the relative scale of generation versus consumption across participants.

6. Results and analysis

This section reports the results obtained by applying the proposed methodology for predicting allocation coefficients in an energy community, namely the Municipal Energy Community (MEC) introduced in Section 5. The implementation is based on supervised machine-learning models, with emphasis on eXtreme Gradient Boosting (XGBoost). Outcomes are evaluated by comparison against the nominal static distribution currently used in the MEC and against alternative dynamic allocation strategies derived within the same learning pipeline. The analysis combines graphical and numerical evidence. Performance is assessed using the indices defined in Section 4.5, which quantify technical efficiency, surplus management, and member-level utilization. The section first presents the XGBoost training and validation results, including the evaluation protocol and error metrics, and then assesses the allocation strategies under realistic operation using annual summaries and time-resolved inspection.

6.1. XGBoost training and validation

As described in Section 4, the predictive models are based on XGBoost, a tree-ensemble method that constructs decision trees sequentially to reduce prediction error. The training stage has two objectives. The first objective is to select suitable hyperparameters under a time-consistent validation protocol. The second objective is to learn the distinct consumption and generation patterns exhibited by the members of the Municipal Energy Community (MEC). For clarity and to ensure consistent notation across datasets and models, each participant is indexed using the tags reported in Table 3.

Table 3
Tag mapping used throughout the methodology to index MEC participants as generators and consumers, ensuring consistent notation across datasets, models, and allocation coefficients.

Assigned Tag	MEC Member
G_1	School P as Generator
G_2	School E as Generator
C_1	School P as Consumer
C_2	School E as Consumer
C_3	Sports Area
C_4	Sports Center
C_5	Civic Center

Model training follows a chronological split to prevent information leakage. The hourly dataset for 2023 is divided into a training segment comprising the first 70% of samples, approximately from January to early August, and a validation segment comprising the remaining 30%, approximately from mid August to December. This protocol reflects deployment conditions, in which coefficient estimates for a future horizon must be produced solely from past observations. Each allocation coefficient is learned using an independent XGBoost regressor so that the mapping can adapt to the variability of each prosumer-consumer pair. Hyperparameters are tuned to balance bias and variance, with particular attention to the learning rate, maximum tree depth, and sub-sampling ratios. The trained models are stored for reuse in the subsequent allocation simulations and in the comparisons against baseline strategies.

During validation, predictions over the final 30% of the year are evaluated against the ideal proportional targets generated offline from historical metering records. Performance is quantified using the mean absolute error, the root mean squared error, and the coefficient of determination, denoted MAE, RMSE, and R^2, respectively. In addition, an energy-weighted MAE is reported by weighting errors with PV generation at each time step. This weighting emphasizes accuracy during PV-rich hours, when coefficient errors have the largest impact on surplus exports and self-consumption under the CSC settlement mechanism.

To assess whether the learned coefficients reproduce the distributional properties of the targets, Fig. 7 reports summary statistics of the

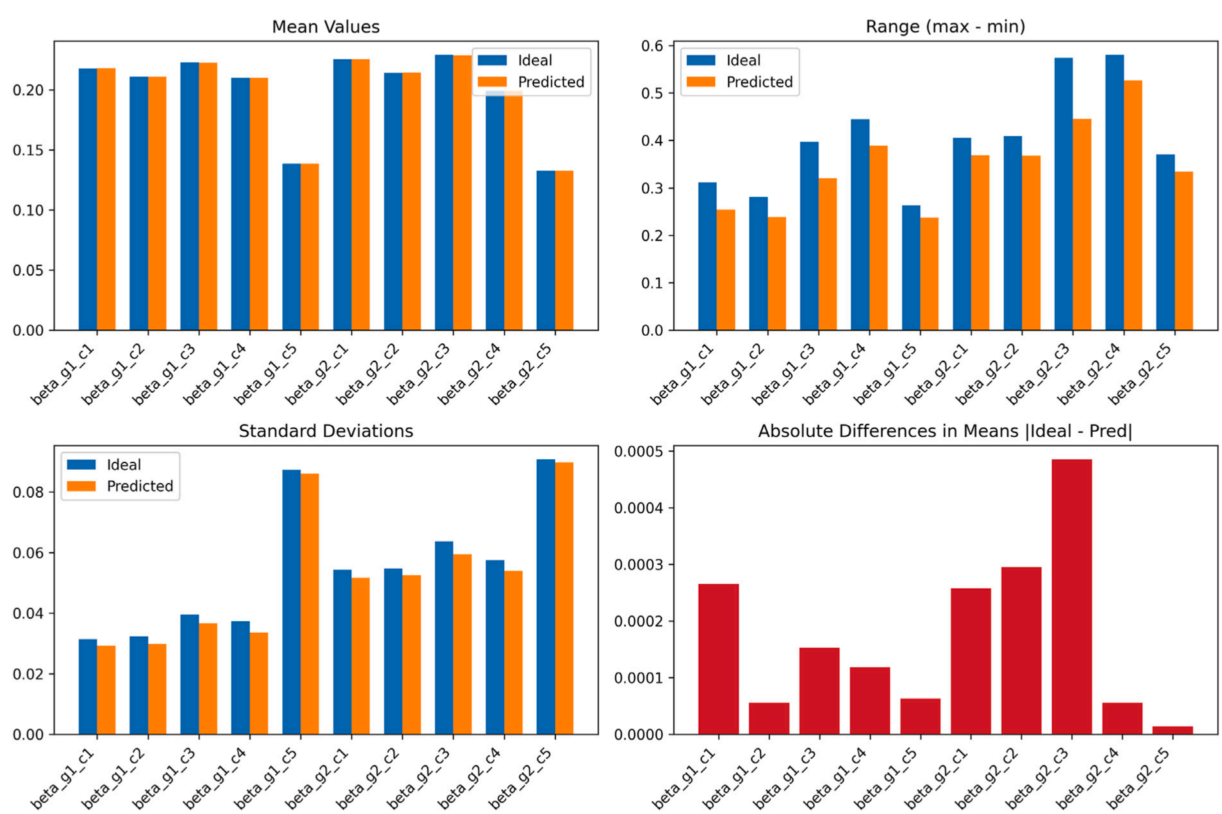


Fig. 7. Statistical comparison of ideal and XGBoost predicted allocation coefficients over the evaluation horizon, reporting mean, range, and standard deviation, together with the absolute deviation between mean values.

ideal and predicted series, including mean values, ranges, and standard deviations, together with the absolute deviation between average values. Preserving these global characteristics is important because persistent shifts in the mean or dispersion would translate into systematic allocation biases.

Complementarily, Fig. 8 reports MAE, RMSE, and R^2 for each coefficient, enabling a coefficient-level assessment of predictive accuracy. Differences across coefficients reflect heterogeneous demand variability and differing allocation relevance across receivers. Nevertheless, the overall consistency of these metrics supports the robustness of the learned mapping across the full coefficient set.

Parity plots for all coefficients are shown in Fig. 9, comparing ideal targets with XGBoost predictions. Concentration of samples around the $y = x$ line, together with high values of R^2, indicates that the models capture both the central tendency and the variability of the target coefficients, which is required to reproduce the intended demand-proportional allocation effect.

Temporal fidelity at the operational scale is illustrated in Fig. 10, which compares predicted and ideal coefficients over a representative week. The predicted coefficients track the dominant intraday variations and respond to short-term changes, supporting their suitability for hour-ahead allocation under CSC operation.

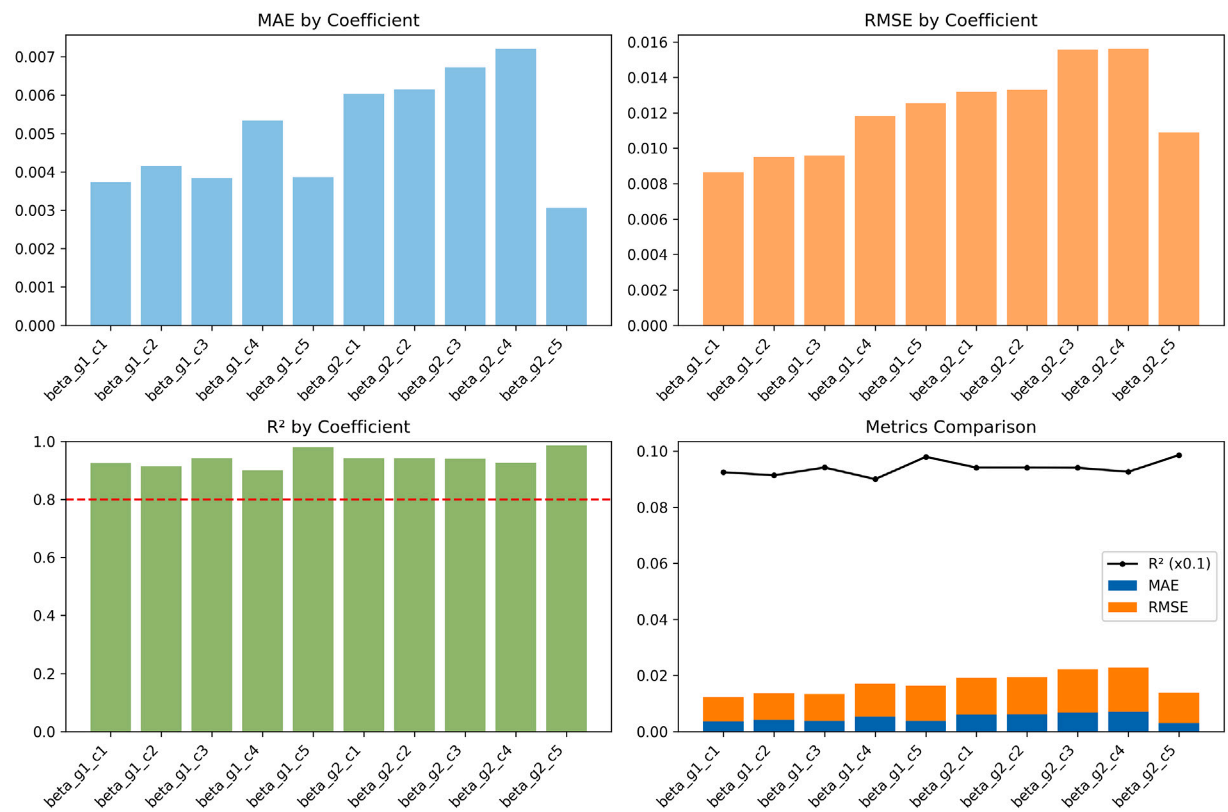


Fig. 8. Per coefficient predictive performance of the XGBoost models over the evaluation horizon, reported using MAE, RMSE, and R^2.

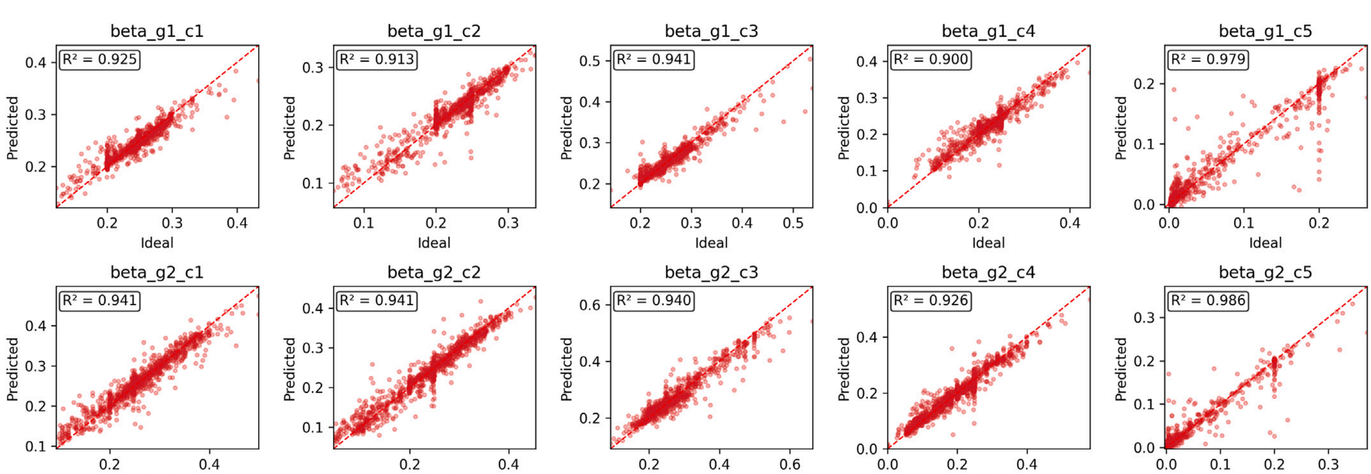


Fig. 9. Parity plots comparing predicted and ideal allocation coefficients for all prosumer-member pairs, illustrating agreement with the ideal targets.

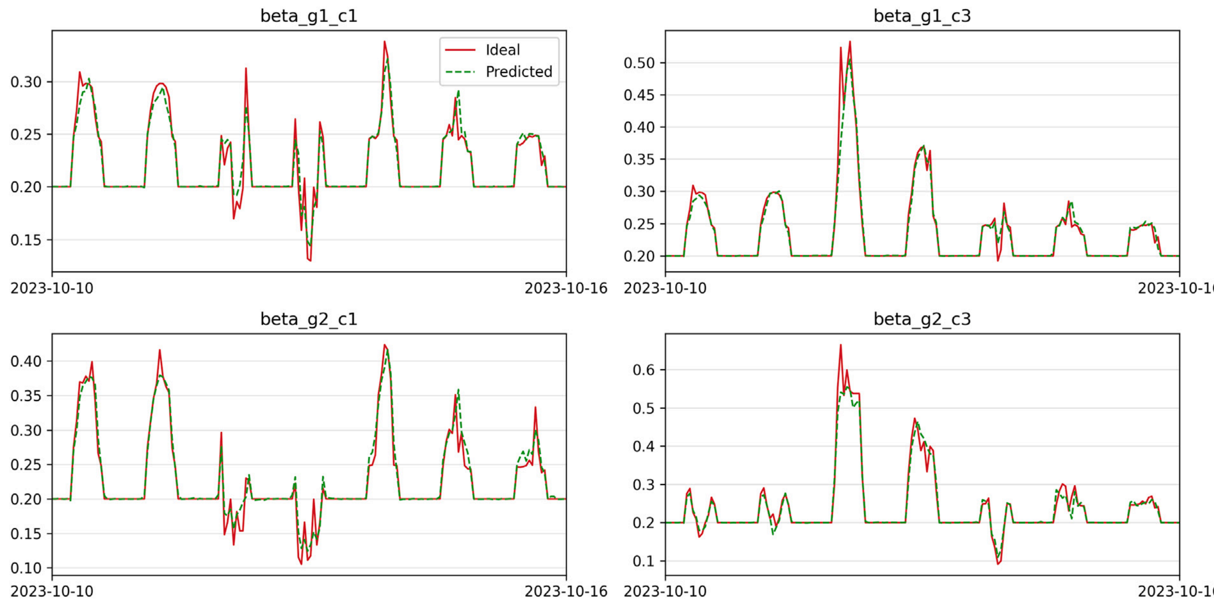


Fig. 10. One-week comparison of ideal and predicted allocation coefficients for selected prosumer-member pairs, illustrating short-term tracking performance relevant for hourly CSC operation.

Table 4
Validation performance of each XGBoost regressor for the allocation coefficients, reported using MAE, RMSE, R^2, and energy-weighted MAE (WMAE).

Target Coefficients	MAE	RMSE	R^2	WMAE Energy
$\hat{\beta}_{g_1,c_1}$	0.0037	0.0086	0.9245	0.0099
$\hat{\beta}_{g_1,c_2}$	0.0042	0.0095	0.9134	0.0101
$\hat{\beta}_{g_1,c_3}$	0.0038	0.0096	0.9415	0.0103
$\hat{\beta}_{g_1,c_4}$	0.0053	0.0118	0.8997	0.0142
$\hat{\beta}_{g_1,c_5}$	0.0039	0.0125	0.9794	0.0057
$\hat{\beta}_{g_2,c_1}$	0.0060	0.0132	0.9410	0.0145
$\hat{\beta}_{g_2,c_2}$	0.0061	0.0133	0.9410	0.0141
$\hat{\beta}_{g_2,c_3}$	0.0067	0.0156	0.9403	0.0170
$\hat{\beta}_{g_2,c_4}$	0.0072	0.0156	0.9262	0.0157
$\hat{\beta}_{g_2,c_5}$	0.0031	0.0109	0.9856	0.0044

The Table 4, summarizes the validation results for each regressor. Across all targets, errors remain small and the explained variance remains high, while the energy-weighted error confirms that accuracy is maintained during high-generation periods.

Before integration into the allocation simulations, the predicted coefficients are post-processed to satisfy the operational constraints of the MEC. In particular, the community includes a citizenship component that receives a fixed share of PV generation at each time step. The coefficient vector is therefore adjusted by normalization so that the resulting allocations remain feasible and consistent with the fixed contractual share.

6.2. Energy community performance analysis

Community performance is assessed by comparing the nominal allocation, which applies the static allocation coefficients currently deployed in the MEC, with the proposed learning-based dynamic strategies. The evaluation employs the technical and economic indicators defined in Section 4.5 and combines annual summaries with time-resolved inspection of representative operating periods. To illustrate the operational mechanisms that drive the aggregate outcomes, a reference week from 10 to 16 August 2023 is analyzed. This interval coincides with the summer holiday period, during which several municipal buildings exhibit atypical occupancy and altered demand profiles. The resulting changes in demand stress the allocation scheme and make mismatches between assigned energy and realized consumption more visible.

6.2.1. Nominal case analysis

In the nominal case, PV generation is distributed according to the static coefficients reported in Table 2. These coefficients remain constant throughout the year and therefore do not reflect day-to-day or seasonal changes in consumption. While this approach is fully compliant with the CSC framework and is straightforward to administer, it can allocate energy to members during hours in which their demand is low, thereby increasing surplus exports. The holiday reference week provides a clear illustration of this limitation. Because the allocation shares are fixed, PV energy continues to be assigned to buildings with reduced activity even when their contemporaneous consumption is near zero. Under CSC settlement rules, any allocated energy that cannot be absorbed by the receiving member is exported, so this systematic over-allocation directly increases the surplus term and reduces effective self-consumption. As a result, part of the community PV production is monetized through the compensation mechanism rather than being used to offset local imports. This behavior weakens the technical objective of maximizing local utilization and limits the associated economic benefit.

The Fig. 11 reports the allocated energy and realized consumption for the five consumer members, together with the corresponding energy balance. Surplus episodes are visible when allocated energy exceeds consumption, which corresponds to negative values of the energy balance as defined in Section 4.5. The Sports Center provides a representative example, as it receives a non-negligible allocated share throughout the week despite exhibiting minimal demand during the same interval. The resulting persistent negative balance indicates that the static coefficients continue to assign PV energy that cannot be utilized locally and is therefore exported. Similar patterns appear in other members with reduced or intermittent activity, confirming that fixed coefficients are unable to accommodate short-term occupancy-driven variations in demand. This nominal operation establishes the baseline against which the dynamic strategies are compared. In particular, it highlights the mechanism through which dynamic coefficients can improve performance, namely by reallocating energy away from low-demand receivers during such intervals and toward members with active consumption, thereby reducing exports and increasing self-consumption without changing total PV production.

6.2.2. Dynamic allocation case analysis

In the dynamic scenario, the MEC is simulated using time-varying allocation coefficients inferred by the proposed XGBoost models. The regressors are trained to approximate the ideal proportional targets introduced in Section 4.2, so that the learned coefficients inherit the intended operational behavior of reallocating PV energy toward members with higher contemporaneous demand. Importantly, coefficient inference relies only on calendar descriptors and lagged metering variables available at decision time, while feasibility is enforced through the normalization step described in Section 4.3. This design allows the allocation policy to adapt at the hourly scale, which is not possible under static allocation coefficients.

The Fig. 12 reports the resulting energy flows for the same August reference week used in the nominal assessment. The mechanism behind the improvement is visible in the redistribution of allocated energy away from low-demand receivers and toward members exhibiting active consumption. In particular, hours in which the Sports Center receives non-negligible allocations under the nominal strategy despite near-zero demand are mitigated under the dynamic coefficients, with a corresponding increase in allocations to the schools and the Sports Area when

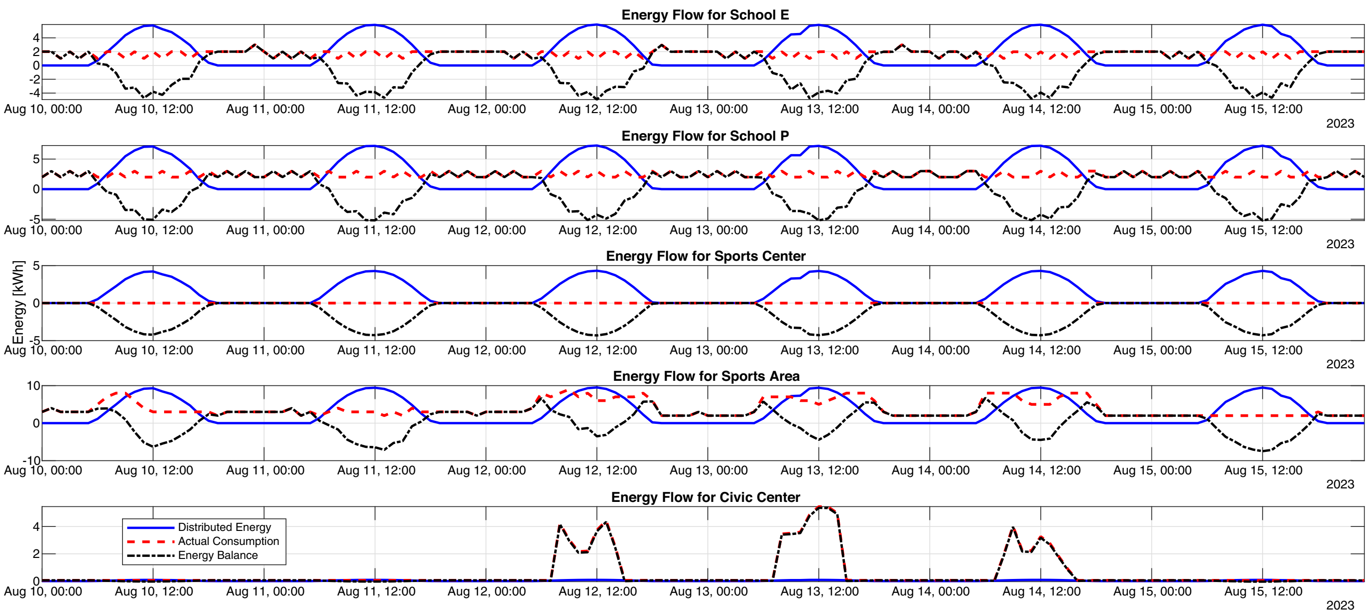


Fig. 11. Member-level time series of allocated energy, realized consumption, and energy balance under nominal static coefficients for the reference week in the MEC.

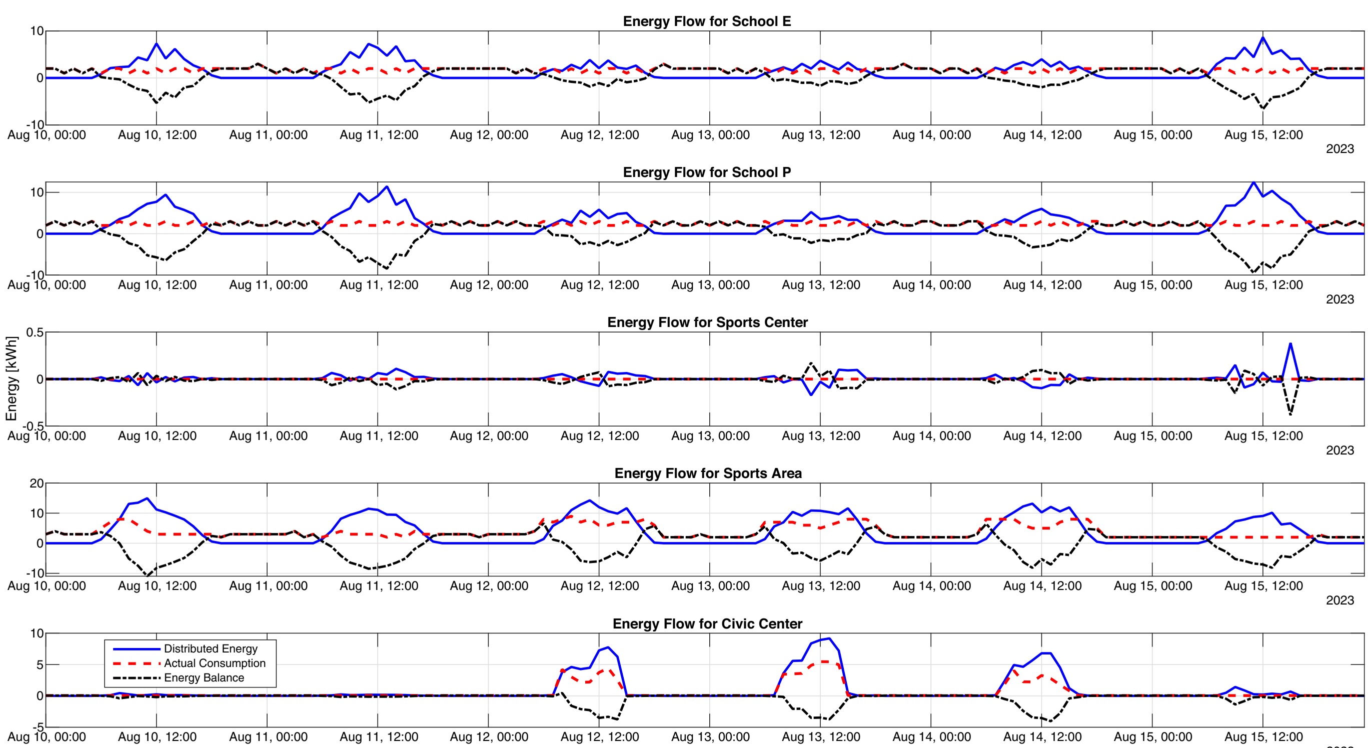


Fig. 12. Member-level time series of allocated energy, realized consumption, and energy balance under XGBoost-based dynamic coefficients for the same representative week in the MEC.

their demand is present. This demand-responsive reallocation reduces the magnitude of negative energy-balance excursions and, consequently, lowers surplus exports while increasing local self-consumption. Overall, the dynamic policy reshapes allocation at the hours where static coefficients systematically over-provision inactive buildings, thereby improving community-level utilization of PV generation under the CSC operational constraints.

6.2.3. Comparative analysis

A comparative evaluation was conducted to benchmark the proposed hourly XGBoost allocation against alternative learning baselines within the same pipeline. Two reduced update policies were considered, namely XGBoost coefficients refreshed every six hours and every twelve hours, to quantify the impact of temporal granularity on allocation effectiveness. In addition, a linear regression baseline was trained independently for each coefficient to isolate the contribution of nonlinear modeling beyond the effect of using dynamic coefficients.

The Fig. 13 reports the mean self-consumption profile by hour of the day over the full-year horizon. Relative to the nominal static allocation, all learning-based strategies increase self-consumption during daylight hours, when PV is available and allocation decisions are active. The improvement is most pronounced for the hourly XGBoost policy, which better tracks intraday variations in building occupancy and demand. By contrast, the six-hour and twelve-hour updates smooth the coefficient trajectories and therefore cannot react to short-lived demand changes, which explains their systematically lower self-consumption gains compared to the hourly model.

The same mechanism is reflected in surplus behavior. The Fig. 14 shows the mean surplus magnitude by hour of the day. Static allocation coefficients generate recurrent surpluses when generation is allocated to

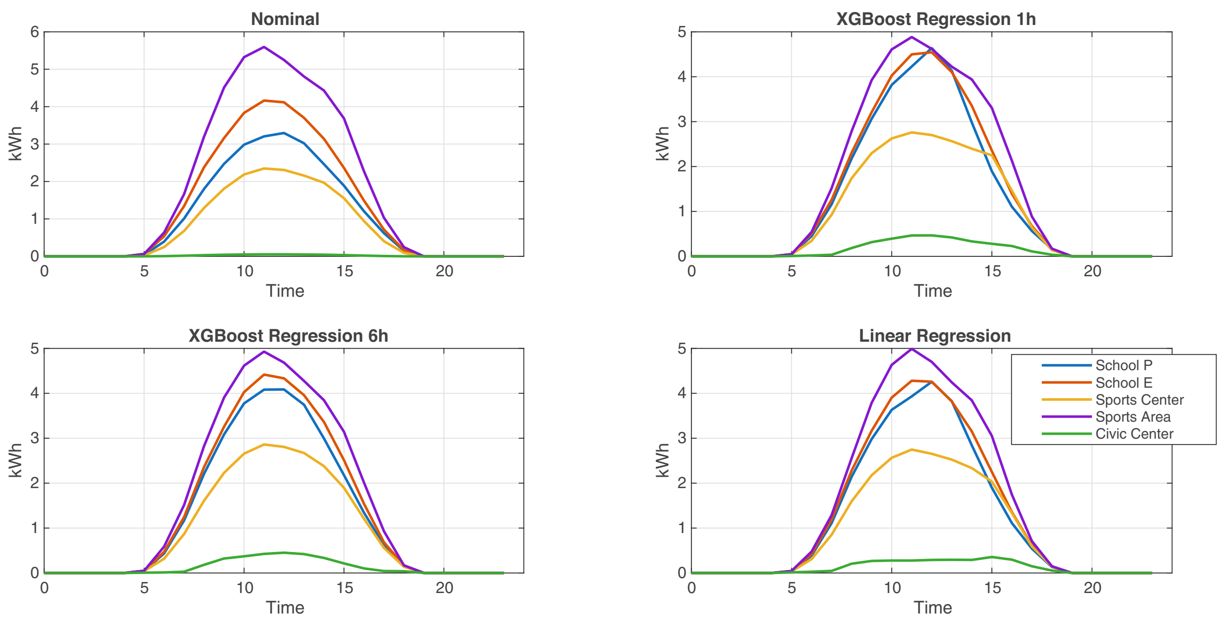


Fig. 13. Average hourly self-consumption profiles of MEC members under the nominal allocation and the learning-based baselines.

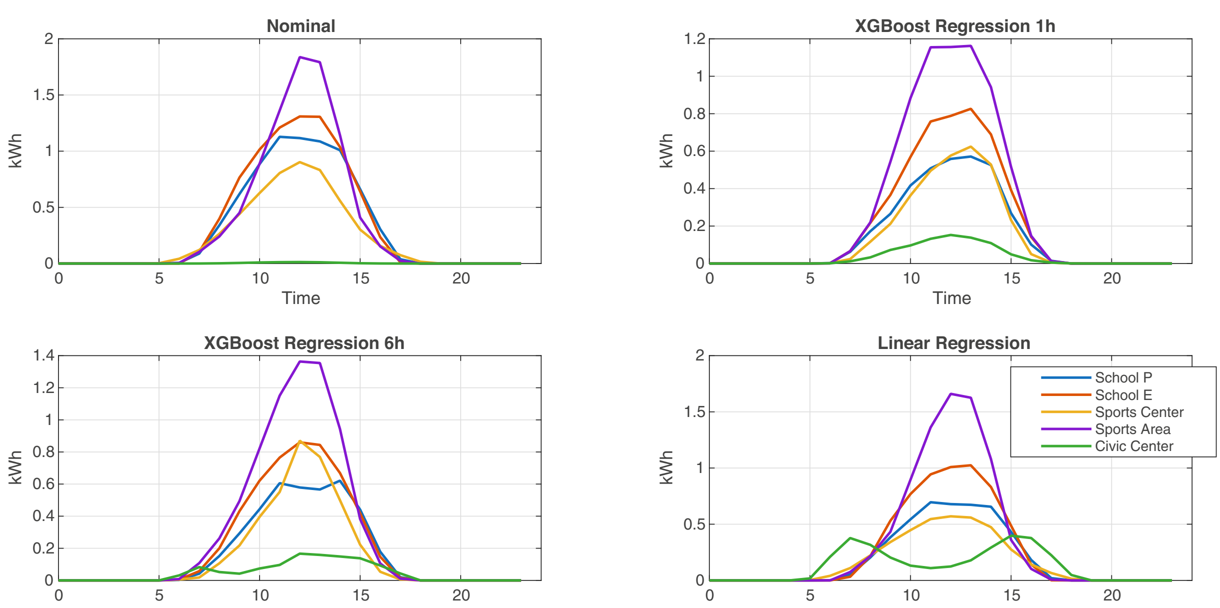


Fig. 14. Average hourly surplus profiles under the nominal allocation and the learning-based baselines.

Table 5
MEC performance metrics values for the allocation strategies.

MEC Metric	Nominal	XGBoost-1h	XGBoost-6h	XGBoost-12h	Linear
Total Consumption (kWh)	212,489.45	212,489.45	212,489.45	212,489.45	212,489.45
Total Distributed (kWh)	53,225.55	53,225.55	53,225.55	53,225.55	53,225.55
Self-Consumption (kWh)	42,664.86	**46,234.03**	45,626.22	45,601.99	44,049.90
Surplus (kWh)	10,560.69	**6991.52**	7599.34	7623.57	9175.65
SFI	0.2581	**0.2419**	0.2421	0.2546	0.3707
ASE	4.67	**3.30**	3.59	3.42	2.83
Surplus Revenue (€)	**713.04**	481.24	518.73	520.05	635.17

Table 6
Energy balance accuracy (EBA) index values for the allocation strategies.

MEC Member	Nominal	XGBoost-1h	XGBoost-6h	XGBoost-12h	Linear
School P	3.51	**3.10**	3.14	3.15	3.21
School E	4.61	**4.42**	4.43	4.43	4.53
Sports Area	3.14	**2.85**	2.90	2.90	2.92
Sports Center	**8.94**	9.08	9.10	9.09	9.18
Civic Center	0.41	**0.32**	0.34	0.35	0.43

members with low contemporaneous demand, a pattern that is exacerbated during periods of reduced activity. Dynamic coefficients mitigate these mismatches by reallocating PV shares toward active consumers. Again, hourly updates achieve the largest reduction because they respond to within-day variations that are averaged out under coarser update intervals.

The Table 5 quantifies the annual effects. Since total consumption and total distributed PV are identical across strategies, performance differences arise exclusively from how distribution is scheduled across members and hours. The proposed XGBoost one-hour policy increases self-consumed energy from 42,664.86 kWh to 46,234.03 kWh, corresponding to an 8.4% gain, and reduces total surplus from 10,560.69 kWh to 6991.52 kWh, corresponding to a 34% reduction. The six-hour and twelve-hour variants retain most of the benefit but exhibit slightly higher surplus and lower self-consumption, consistent with the reduced ability to react to short-term demand changes. The linear baseline yields only modest improvements, which indicates that capturing nonlinear interactions between calendar structure and lagged demand and generation is important for effective coefficient prediction.

Table 5 also highlights an economic implication. Surplus revenue decreases under the XGBoost policies because a smaller fraction of PV is exported. This result is consistent with the objective of maximizing local self-consumption, and it illustrates that exported surplus has limited economic value under the considered compensation mechanism. Consequently, the proposed allocation prioritizes technical utilization and demand coverage rather than export monetization.

Member-level indices provide further insight into the redistribution mechanisms. The Table 6 shows that the hourly XGBoost policy improves energy balance accuracy for most members, indicating tighter matching between allocated energy and realized demand. Table 7 shows that receiver utilization increases substantially for the main demand drivers, which confirms that the dynamic policy reallocates energy away from hours and receivers where it would remain unused. Table 8 shows reduced grid dependency for most members, meaning that a larger share of demand is met through community PV sharing. The exceptions, such as the Sports Center, are consistent with the reference-week evidence: when a member exhibits extended low-demand periods, a demand-responsive policy reallocates energy toward receivers with a higher probability of immediate utilization. This behavior improves community-level efficiency but can slightly reduce demand matching for specific receivers in some intervals, motivating future inclusion of fairness constraints or minimum coverage guarantees.

Finally, the community-aggregated comparison in Fig. 15 summarizes these trade-offs across strategies. The hourly XGBoost policy provides the most balanced outcome across utilization, demand matching, and grid dependency, which is consistent with its superior surplus reduction and self-consumption gains observed in Table 5.

Table 7
Receiver utilization index (RUI) values for the allocation strategies.

MEC Member	Nominal	XGBoost-1h	XGBoost-6h	XGBoost-12h	Linear
School P	0.77	**0.90**	0.88	0.88	0.85
School E	0.80	**0.87**	0.87	0.86	0.84
Sports Area	0.78	**0.88**	0.86	0.86	0.84
Sports Center	0.84	**0.85**	0.84	0.84	0.80
Civic Center	**0.87**	0.80	0.72	0.71	0.49

Table 8
Grid dependency ratio (GDR) index values for the allocation strategies.

MEC Member	Nominal	XGBoost-1h	XGBoost-6h	XGBoost-12h	Linear
School P	0.76	**0.70**	0.70	0.70	0.72
School E	0.77	**0.76**	0.76	0.76	0.77
Sports Area	0.80	**0.74**	0.75	0.75	0.76
Sports Center	**0.83**	0.84	0.85	0.86	0.86
Civic Center	0.96	**0.67**	0.71	0.72	0.73

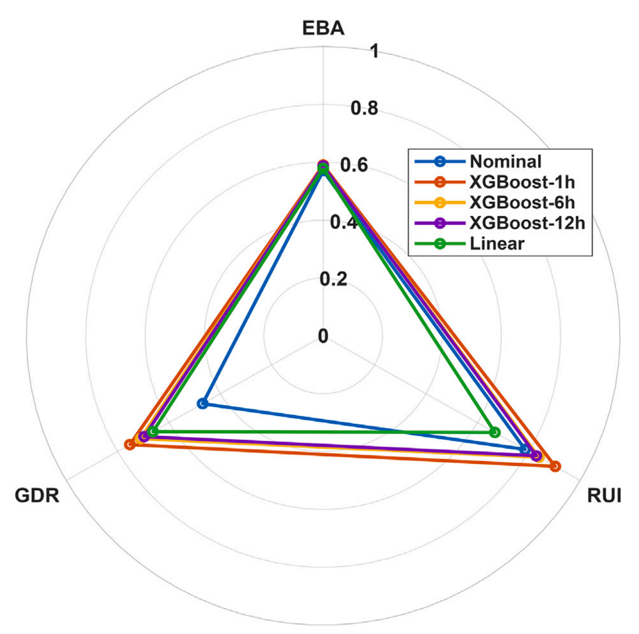


Fig. 15. Community-level radar plot summarizing normalized average values of EBA, RUI, and GDR for each allocation strategy.

7. Conclusions and future work

This work introduces and validates a data-driven methodology to compute time-varying allocation coefficients for Collective Self-Consumption based energy communities, using a real Municipal Energy Community in Catalonia as a case study. The proposed approach combines an ideal proportional allocation rule, used offline to generate supervised learning targets, with XGBoost regressors that infer deployable coefficients from calendar descriptors and lagged metering variables. The resulting allocation policy is subsequently post-processed to satisfy the CSC conservation constraint, thereby ensuring operational feasibility.

The results confirm that dynamic coefficient prediction can materially improve the utilization of locally generated PV energy under the CSC settlement structure. Relative to the nominal static allocation, hourly XGBoost updates increased community self-consumption from 42,664.86 kWh to 46,234.03 kWh, corresponding to an 8.4% improvement, while reducing surplus injections from 10,560.69 kWh to 6,991.52 kWh, corresponding to a 34% reduction. This reduction is consistent across complementary indicators, with the Surplus Frequency Index decreasing from 0.258 to 0.242 and the Average Surplus per Event decreasing from 4.67 kWh to 3.30 kWh. Although surplus revenues decreased, the observed shift reflects the intended operational outcome of CSC-based energy communities, namely prioritizing local self-consumption and reducing avoidable exports.

From a practical implementation perspective, the proposed framework is compatible with the information flows typically available in CSC deployments. It requires only standard hourly metering records of member consumption and prosumer PV generation, together with timestamps from which calendar features can be derived, and it does not rely on intrusive measurements such as occupancy or user-specific behavioral data. Inference is lightweight and can be executed at community scale, since each coefficient is obtained from a regression model evaluated on a low-dimensional feature vector and feasibility is ensured through a simple normalization step. This makes the methodology suitable for integration into aggregator, retailer, or community energy management platforms that periodically publish allocation coefficient schedules.

The approach is also policy-compatible in the sense that it operates entirely within the coefficient-based allocation mechanism of the Spanish CSC framework. Coefficients are produced to satisfy conservation at each hour and can be exported as a time series suitable for reporting and settlement. The results further highlight a practical implication for current compensation regimes. Because exported surplus is monetized under time-varying and contract-dependent prices, reducing surplus is not only an efficiency objective but also a risk-mitigation strategy against unfavorable compensation intervals. In this context, dynamic allocation provides an actionable lever to improve local utilization without requiring changes to physical infrastructure.

Several limitations motivate future work. First, the present evaluation assumes complete and reliable hourly metering data, and extending the pipeline to explicitly handle missing data, noisy measurements, and partial observability will be important for broader applicability. Second, validation was conducted on a single MEC, and additional case studies across different community typologies, including citizen energy communities and hybrid configurations, are required to assess generalization and robustness. Third, while XGBoost provided a strong accuracy-complexity trade-off, alternative temporal learning architectures such as recurrent networks, temporal convolutional models, or attention-based transformers may be explored to better capture long-range seasonality and regime shifts.

Beyond model improvements, future research should consider decision objectives that reflect broader community priorities. Extensions could incorporate multi-objective formulations that balance self-consumption maximization, economic performance under compensation uncertainty, and equity-oriented criteria across members. In addition, periodic retraining strategies, uncertainty quantification, and explainability analyses can strengthen transparency and facilitate stakeholder acceptance. Finally, practical deployment would benefit from studying operational procedures for coefficient publication and updating, including the interaction between coefficient update frequency, administrative constraints, and the stability of member-level outcomes.

In summary, the proposed methodology demonstrates that learning-based dynamic coefficients can substantially improve CSC community operation while remaining compatible with the prevailing settlement mechanism. The framework increases self-consumption by 8.4% and reduces surplus injections by 34% in the considered MEC, establishing a scalable and deployable pathway to replace static allocation with adaptive, data-driven energy sharing. These results support the view that integrating machine learning into CSC allocation processes is a practical step toward more efficient and resilient energy communities under existing regulatory conditions.

CRediT authorship contribution statement

Sebastián Madrigal: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Ramon Gallinad:** Visualization, Software, Data curation. **Jose L. Vicario:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis. **Antoni Morell:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Ramon Vilanova:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis.

Declaration of competing interest

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

Ramon Vilanova and Jose Lopez Vicario report that financial support was provided by the Catalan Government, the Spanish Ministry of Science and Innovation and the European Union. 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.

Acknowledgements

This work has received support from the Catalan Government under Project 2021-SGR-00197 and also by the Spanish Government under MICINN projects PID2024-156522OB-C33 co-funded with the European Union ERDF funds, and projects OPEN4CEC PCI2023-145975-2 and SMART-LEM PCI2023-145977-2 funded by MCIU/AEI/10.13039/501100011033 and co-funded by the European Union. We also express our gratitude to the Barcelona Provincial Council, in particular to the Technical Office for Climate Change and Sustainability, for providing the data employed in the case study analyzed in this work.

Appendix A. Nomenclature and symbols

A.1. Sets, indices, and dimensions

The sets, indices, and dimensions are reported in Table A.9.

Table A.9
Sets, indices, and dimensions used throughout the paper.

Symbol	Definition
$\mathcal{C} = \{c_1, \ldots, c_K\}$	Set of consumers in the CSC community
$\mathcal{P} = \{p_1, \ldots, p_I\}$	Set of prosumers in the CSC community
$\mathcal{N} = \mathcal{C} \cup \mathcal{P}$	Set of all members in the community
K	Number of consumers
I	Number of prosumers
$N = \|\mathcal{N}\| = I + K$	Total number of members
n	Generic member index, $n \in \mathcal{N}$
k	Consumer index, $k \in \mathcal{C}$
i	Prosumer index, $i \in \mathcal{P}$
t	Time index
T	Number of time steps in the evaluation horizon
$B = I \cdot N$	Number of allocation coefficients under the prosumer-centered definition

A.2. Energy variables and CSC allocation

The energy variables and CSC allocation quantities adopted in this work are listed in Table A.10.

Table A.10
Energy variables and CSC allocation quantities.

Symbol	Definition
$C_n(t)$	Electricity consumption of member n at time t
$C_k(t)$	Electricity consumption of consumer c_k at time t
$C_i(t)$	Electricity consumption of prosumer p_i at time t
$\mathbf{C}(t) = [C_1(t), \ldots, C_N(t)]$	Vector of member consumptions at time t
$G_i(t)$	PV generation of prosumer p_i at time t
$G_{\text{tot}}(t) = \sum_{i=1}^{I} G_i(t)$	Total PV generation of the community at time t
$\text{TEG}(t)$	Total Energy Generated for CSC at time t
$\text{INE}_n(t)$	Individualized Net Energy assigned to member n at time t
α	Self-consumption fraction applied to PV generation under CSC (community-specific)
$\beta_n(t)$	Allocation coefficient for member n at time t (single-source formulation)
$\beta_{p_i,n}(t)$	Allocation coefficient assigning prosumer p_i generation to member n at time t
$\beta^*_{p_i,n}(t)$	Ideal proportional allocation coefficient used as supervised-learning target
$\hat{\beta}_{p_i,n}(t)$	Raw XGBoost prediction of the allocation coefficient
$\tilde{\beta}_{p_i,n}(t)$	Post-processed and normalized allocation coefficient used in operation
$D_n(t)$	Energy allocated to member n at time t through CSC coefficients

A.3. Learning variables and feature definitions

The machine-learning notation and feature definitions used in the paper are provided in Table A.11.

Table A.11
Machine learning notation and feature definitions.

Symbol	Definition
$f_\theta(\cdot)$	Learned mapping implemented by an XGBoost regressor with parameters θ
$x_n^{\text{cal}}(t)$	Calendar feature vector for member n at time t
$x_n^{\text{lags}}(t)$	Lagged consumption feature vector for member n at time t
$x_{p_i}^{\text{lags}}(t)$	Lagged PV generation feature vector for prosumer p_i at time t
θ	Model parameter set of a trained regressor

A.4. Heuristic target construction variables

The variables used in the ideal proportional heuristic for constructing supervised-learning targets are reported in Table A.12.

Table A.12
Variables used in the ideal proportional heuristic employed to generate supervised-learning targets.

Symbol	Definition
$d_n(t) = \min\big(C_n(t), G_{\text{tot}}(t)\big)$	Effective demand of member n at time t
$D(t) = \sum_{m=1}^{N} d_m(t)$	Aggregate effective demand at time t
$\sum_{n=1}^{N} \beta^*_{p_i,n}(t) = 1$	CSC conservation constraint satisfied by the ideal coefficients for each prosumer

A.5. Performance metrics and economic variables

The performance metrics and economic variables used in the evaluation are summarized in Table A.13.

Table A.13
Performance metrics and economic variables used in the evaluation.

Symbol	Definition
$\text{ES}(t)$	Energy surplus exported at time t
$\text{EB}_n(t) = C_n(t) - D_n(t)$	Energy balance of member n at time t
SFI	Surplus Frequency Index
ASE	Average Surplus per Event
RUI_n	Receiver Utilization Index of member n
GDR_n	Grid Dependency Ratio of member n
EBA_n	Energy Balance Accuracy of member n
ω_t	PV availability weight at time t used in WMAE
$\pi_{\text{comp}}(t)$	Surplus compensation price at time t
Surplus revenue	Monetary revenue from exported surplus valued by $\pi_{\text{comp}}(t)$

A.6. Acronyms

The acronyms used in the paper are compiled in Table A.14.

Table A.14
Acronyms used in the paper.

Acronym	Definition
ASE	Average Surplus per Event
CSC	Collective Self-Consumption
DSO	Distribution System Operator
EBA	Energy Balance Accuracy
EB	Energy Balance
EC	Energy Community
ES	Energy Surplus
GDR	Grid Dependency Ratio
INE	Individualized Net Energy
MAE	Mean Absolute Error
MEC	Municipal Energy Community
ML	Machine Learning
PV	Photovoltaic
RMSE	Root Mean Squared Error
RUI	Receiver Utilization Index
SFI	Surplus Frequency Index
TEG	Total Energy Generated
WMAE	Energy-weighted Mean Absolute Error
XGBoost	eXtreme Gradient Boosting

Data availability

The authors are unable or have chosen not to specify which data have been used.

References

[1] Iazzolino G, Sorrentino N, Menniti D, Pinnarelli A, De Carolis M, Mendicino L. Energy communities and key features emerged from business models review. Energy Policy 2022;165:112929. https://doi.org/10.1016/j.enpol.2022.112929
[2] Ahmed S, Ali A, D'Angola A. A review of renewable energy communities: concepts, scope, progress, challenges, and recommendations. Sustainability 2024;16(5):1749. https://doi.org/10.3390/su16051749
[3] Rumbayan M, Kindangen J, Sambul A, Sompie S, Cross J. Solar energy implementation in rural communities and its contributions to SDGs: a systematic literature review. Unconv Resour 2025:100180. https://doi.org/10.1016/j.uncres.2025.100180
[4] Bjarghov S, Loschenbrand M, Saif AUNI, Pedrero RA, Pfeiffer C, Khadem SK, Rabelhofer M, Revheim F, Farahmand H. Developments and challenges in local electricity markets: a comprehensive review. IEEE Access 2021;9:58910–43. https://doi.org/10.1109/access.2021.3071830

[5] Araújo I, Grasel B, Cerveira A, Baptista J. Energy sharing models in renewable energy communities. In: 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET); 2023. p. 1–6. https://doi.org/10.1109/icecet58911.2023.10389585
[6] Palou JTV, Gonzalez JS, Santos JMR, Alonso CA, Fernandez JMR. Sharing approaches in collective self-consumption systems: a techno-economic analysis of the spanish regulatory framework. Energy Strateg Rev 2023;45:101055. https://doi.org/10.1016/j.esr.2023.101055
[7] Mena AJG, Medina VFN, Bouakkaz A, Haddad S. Analysis and optimisation of collective self-consumption in residential buildings in Spain. Energy and Buildings 2023;283:112812. https://doi.org/10.1016/j.enbuild.2023.112812
[8] Camblong H, Curea O, Ugartemendia J, Boussaada Z, Lizarralde I, Etxegarai G. Photovoltaic energy sharing: implementation and tests on a real collective self-consumption system. Heliyon 2023;9(12):e22252. https://doi.org/10.1016/j.heliyon.2023.e22252
[9] Casella V, Ferro G, Parodi L, Robba M. Maximizing shared benefits in renewable energy communities: a bilevel optimization model. Appl Energy 2025;386:125562. https://doi.org/10.1016/j.apenergy.2025.125562
[10] Hamze FJS, Guerrero-Lemus R. Development and evaluation of an optimization model for enhancing energy communities sharing coefficients in Spain. Energy and Buildings 2025:115951. https://doi.org/10.1016/j.enbuild.2025.115951
[11] Hernandez-Matheus A, Löschenbrand M, Berg K, Fuchs I, Aragüés-Peñalba M, Bullich-Massagué E, Sumper A. A systematic review of machine learning techniques related to local energy communities. Renew Sustain Energy Rev 2022;170:112651. https://doi.org/10.1016/j.rser.2022.112651
[12] Zhou Y. Advances of machine learning in multi-energy district communities- mechanisms, applications and perspectives. Energy and AI 2022;10:100187. https://doi.org/10.1016/j.egyai.2022.100187
[13] Frieden D, Tuerk A, Roberts J, D'Herbemont S, Gubina AF, Komel B. Overview of emerging regulatory frameworks on collective self-consumption and energy communities in Europe. In: 2022 18th International Conference on the European Energy Market (EEM); 2019. p. 1–6. https://doi.org/10.1109/eem.2019.8916222
[14] Moreno A, Mello J, Villar J. Comparison of selected self-consumption regulatory approaches in Europe. Heliyon 2025;11(12):e43534. https://doi.org/10.1016/j.heliyon.2025.e43534
[15] Madrigal S, Vicario JL, Morell A, Vilanova R. Enhancement review of the collective Self-Consumption based energy communities: insights from a spain-based case. In: 2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA); 2025. p. 1–8. https://doi.org/10.1109/ETFA65518.2025.11205679
[16] Gallego-Castillo C, Heleno M, Victoria M. Self-consumption for energy communities in spain: a regional analysis under the new legal framework. Energy Policy 2021;150:112144. https://doi.org/10.1016/j.enpol.2021.112144
[17] Dasí-Crespo D, Roldán-Blay C, Escrivá-Escrivá G, Roldán-Porta C. Evaluation of the spanish regulation on self-consumption photovoltaic installations. A case study based on a rural municipality in Spain. Renewable Energy 2023;204:788–802. https://doi.org/10.1016/j.renene.2023.01.055
[18] Casella V, Ferro G, Parodi L, Robba M. Energy community optimal management: a bilevel approach. In: 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE); 2024. p. 2720–5. https://doi.org/10.1109/case59546.2024.10711358
[19] Casella V, Farina L, Ferro G, Parodi L, Robba M. Operational management of multiple energy communities in the energy market: a bilevel optimization-based approach. IFAC-PapersOnLine 2025;59(9):133–8. https://doi.org/10.1016/j.ifacol.2025.08.125
[20] Tasmant H, Bossoufi B, Alaoui C, Siano P. A review of machine learning and IoT-based energy management systems for AC microgrids. Comput Electr Eng 2025;127:110563. https://doi.org/10.1016/j.compeleceng.2025.110563
[21] Elomari Y, Mateu C, Marín-Genescà M, Boer D. A data-driven framework for designing a renewable energy community based on the integration of machine learning model with life cycle assessment and life cycle cost parameters. Appl Energy 2024;358:122619. https://doi.org/10.1016/j.apenergy.2024.122619
[22] Piras G, Muzi F, Ziran Z. Open tool for automated development of renewable energy communities: artificial intelligence and machine learning techniques for methodological approach. Energies 2024;17(22):5726. https://doi.org/10.3390/en17225726
[23] Matos M, Almeida J, Gonçalves P, Baldo F, Braz FJ, Bartolomeu PC. A machine Learning-Based electricity consumption forecast and management system for renewable energy communities. Energies 2024;17(3):630. https://doi.org/10.3390/en17030630
[24] Boumaiza A, Maher K. An adaptive home energy management system for prosumers in peer-to-peer trading networks with machine learning optimization. Energy Strateg Rev 2025;60:101749. https://doi.org/10.1016/j.esr.2025.101749
[25] Guiducci L, Palma G, Stentati M, Rizzo A, Paoletti S. A reinforcement learning approach to the management of renewable energy communities. In: 2022 11th Mediterranean Conference on Embedded Computing (MECO); 2023. p. 1–8. https://doi.org/10.1109/meco58584.2023.10154979
[26] Palma G, Guiducci L, Stentati M, Rizzo A, Paoletti S. Reinforcement learning for energy community management: a European-Scale study. Energies 2024;17(5):1249. https://doi.org/10.3390/en17051249
[27] Gonçalves R, Gomes D, Antunes M. Intelligent management of renewable energy communities: an MLaaS framework with RL-based decision making. Energies 2025;18(13):3477. https://doi.org/10.3390/en18133477
[28] Barabino E, Fioriti D, Guerrazzi E, Mariuzzo I, Poli D, Raugi M, Razaei E, Schito E, Thomopulos D. Energy communities: a review on trends, energy system modelling, business models, and optimisation objectives. Sustain Energy Grids Netw 2023;36:101187. https://doi.org/10.1016/j.segan.2023.101187
[29] Government of Spain. Royal Decree-Law 15/2018, of October 5, on urgent measures for the energy transition and consumer protection, Ministry of the Presidency, Parliamentary Relations and Equality, 2018. https://www.boe.es/eli/es/rdl/2018/10/05/15
[30] Government of Spain. Royal decree 244/2019, of April 5, regulating the administrative, technical and economic conditions for self-consumption of electrical energy, Ministry of the Presidency, Parliamentary Relations and Equality, 2019. https://www.boe.es/eli/es/rd/2019/04/05/244
[31] Government of Spain. Order TED/1247/2021, of November 15, which modifies, for the implementation of variable distribution coefficients in collective self-consumption, annex i of royal decree 244/2019, of April 5, BOE, núm.274, november 16 of 2021, pp. 141114–141119, ministry for Ecological Transition and the Demographic Challenge modifies Annex I of Royal Decree 244/2019 to allow hourly variable distribution coefficients, reviewable at least every four months. Nov, 2021.
[32] Institute for Energy Diversification and Saving (IDAE). Collective self-consumption guide; 2024.
[33] Abbasimehr H, Paki R, Bahrini A. A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. Sustain Comput Inform Syst 2023;38:100863. https://doi.org/10.1016/j.suscom.2023.100863
[34] Xu Y, Zheng S, Zhu Q, Wong K-C, Wang X, Lin Q. A complementary fused method using GRU and XGBoost models for long-term solar energy hourly forecasting. Expert Syst Appl 2024;254:124286. https://doi.org/10.1016/j.eswa.2024.124286
[35] Gallinad R, Madrigal S, Vicario JL, Morell A, Vilanova R. Optimizing energy allocation with the CEC-UAB simulator: forecasting comparison. In: 2025 IEEE 30th International Conference on Emerging Technologies and Factory Automation (ETFA); 2025. p. 1–7. https://doi.org/10.1109/ETFA65518.2025.11205651
[36] ESIOS. Spanish electricity grid compensation prices for collective Self-Consumption, Online, available: 2025. https://www.esios.ree.es/es
[37] Climate-Consulting. General pricing for solar panel installations in 2024 Oct 2023. https://climate.selectra.com/es/placas-solares/instalacion/precio