Applied Energy 426 (2026) 128776

Contents lists available at ScienceDirect

Applied Energy

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

Regulation-aware combinatorial optimization of energy community formation under geographic proximity constraints

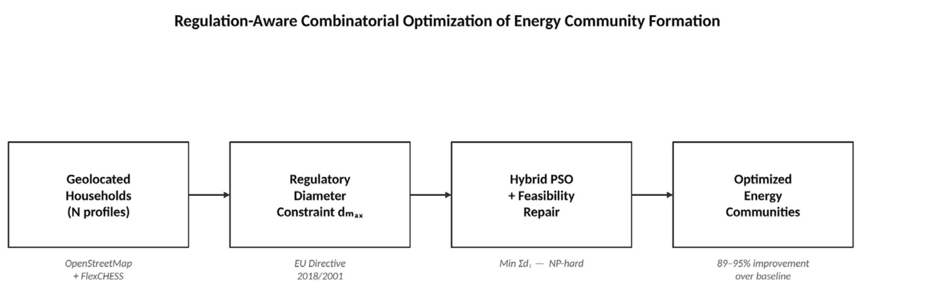
Yacine Merrad, Seifeddine Ben Elghali*

Laboratoire d'Informatique et Systèmes (LIS), UMR CNRS 7020, Université Aix-Marseille, Marseille, 13013, France

HIGHLIGHTS

- REC formation formalized as a diameter-constrained NP-hard partitioning problem under European proximity rules.
- Hybrid PSO with feasibility repair and memetic refinement minimizes annual imbalance across bounded communities.
- Quadratic imbalance minimization proven to tighten the annual self-consumption upper bound.
- 86–93% reduction in imbalance over random and geographic baselines; 7–9% over greedy search (Wilcoxon $p < 0.004$).
- Non-circular hourly validation confirms the annual proxy (rho = −0.42); robust from 1000 to 1900 kWh/kWp.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Renewable energy communities
Collective self-consumption
Geographically constrained partitioning
Annual balance complementarity
Particle swarm optimization
NP-hardness
Regulation-aware design

ABSTRACT

Renewable Energy Communities (RECs) enable collective self-consumption by allowing geographically proximate households to share locally generated renewable energy. While most existing studies optimize the operation of a pre-defined community, the prior problem of how to partition a set of households into multiple regulation-compliant communities remains underexplored. This paper formalizes REC formation as a pre-operational combinatorial design problem: households are partitioned under a strict diameter-based proximity constraint derived from European national regulatory rules. Within this feasible space, we minimize the quadratic imbalance of yearly community energy balances, a structural proxy for self-consumption potential equivalent to variance minimization over community balances, and prove it tightens the theoretical upper bound of annual aggregate self-consumption. The problem is shown to be NP-hard via reduction from 3-PARTITION. A hybrid particle swarm optimization (PSO) framework with strict feasibility repair and local memetic refinement is proposed. Experiments on residential data from two urban zones of Murcia, Spain ($N = 200$, $N = 151$ households; 130 profiles from the FlexCHESS European project) under the Spanish regulatory distance cap of 2 km (Real Decreto-ley 29/2021) demonstrate 86–93% reduction in quadratic imbalance over random and geographic baselines, and 7–9% over greedy local search (Wilcoxon $p < 0.004$, 10 independent runs). Hourly validation on independently generated PVGIS-calibrated profiles (annual balances derived *a posteriori*, eliminating reconstruction circularity) reveals a strong negative correlation between annual imbalance and hourly self-consumption ($\rho = -0.42$, $p < 10^{-5}$, $\rho^2 \approx 0.18$), supporting the annual proxy as a useful pre-operational structural criterion. Ablation

* Corresponding author.
Email addresses: yacine.merrad@lis-lab.fr (Y. Merrad), seifeddine.benelghali@lis-lab.fr (S. Ben Elghali).

https://doi.org/10.1016/j.apenergy.2026.128776
Received 14 March 2026; Received in revised form 5 July 2026; Accepted 28 August 2026

studies confirm robustness to PSO hyperparameters, and climate sensitivity analysis shows PSO dominance across irradiance levels from 1000 to 1900 kWh/kWp. Beyond solution quality, the proposed framework is intended as a pre-operational planning and screening tool that can assist municipalities, community organizers, and Distribution System Operators in identifying promising community configurations before undertaking detailed network-constrained or operational studies.

1. Introduction

Renewable Energy Communities (RECs) are a legally recognized framework in which households, businesses, and public entities jointly own, manage, and share renewable generation and flexibility resources. In the European Union, RECs and Citizen Energy Communities (CECs) are governed by Directive (EU) 2018/2001 (RED II) [1] and Directive (EU) 2019/944 (IEMD) [2], respectively. A fundamental economic asymmetry motivates community participation: the electricity purchase price π^{buy} incorporates network tariffs and taxes, while surplus exported to the grid is remunerated at the lower wholesale rate π^{sell}, with $\pi^{\text{buy}}/\pi^{\text{sell}}$ ranging from two to four in many European markets [3,4]. Internal matching at an intermediate settlement price $\lambda \in [\pi^{\text{sell}}, \pi^{\text{buy}}]$ is therefore Pareto-improving for all participants, independently of subsidies.

Despite this motivation, the research literature has developed along two separate trajectories: the first optimizes the *operation* of a community with fixed membership [3,5–7]; the second studies socio-institutional dynamics of community formation [8,9]. Large-scale sensitivity studies confirm that performance is highly sensitive to membership characteristics—size, prosumer share, load profile diversity—and that composition can dominate the effect of the chosen management strategy [10,11]. Yet the *formation problem*—how to algorithmically partition a set of households into multiple locally bounded communities—has received little attention as a constrained combinatorial optimization problem. From a practical perspective, community formation constitutes an early-stage planning problem that typically precedes detailed grid studies. Municipalities, energy agencies, community organizers, and Distribution System Operators frequently need to identify candidate community configurations before network reinforcement assessments, power-flow analyses, or operational optimization studies can be conducted. The proposed framework addresses this planning stage by constructing geographically feasible and energetically balanced communities using information that is generally available at the outset of a project, namely geographic locations and annual energy balances. This gap is practically significant: national regulations impose strict geographic eligibility rules that constrain the feasible partition space. Multiple legally valid partitions coexist yet lead to markedly different levels of internal matching. To the best of our knowledge, no prior work addresses multi-community formation under a strict diameter-based regulatory constraint with variable induced cardinality K.

Contributions

1. We formalize REC formation as a diameter-constrained partitioning problem with variable cardinality K and establish NP-hardness by reduction from 3-PARTITION [12].
2. We adopt the quadratic imbalance $J(C) = \sum_k S_k^2$, justify it against four competing objectives, and prove it tightens the annual self-consumption upper bound (Corollary 1).
3. We develop a hybrid PSO framework with strict feasibility repair and local memetic refinement enforcing regulatory compliance, treating K as an induced variable.
4. We validate on Murcia (Spain) residential data ($N = 200$, $N = 151$; FlexCHESS profiles) under the Spanish 2 km regulatory cap: 86–93% improvement over random and geographic baselines, 7–9% over greedy search (Wilcoxon $p < 0.004$, 10 runs), robust across $d_{\max} \in \{0.5, 1.0, 2.0, 5.0\}$ km.
5. We validate the annual proxy via non-circular hourly simulation (PVGIS-calibrated profiles with balances derived a posteriori), and position the framework as a pre-operational screening tool for identifying promising REC configurations prior to detailed network and operational studies.

The paper is organized as follows. Section 2 reviews related work. Section 3 presents the regulatory and market framework. Section 4 formalizes the problem. Section 5 presents the PSO framework. Section 6 provides theoretical guarantees. Section 7 presents experimental results. Section 8 gives the exact MILP benchmark. Section 9 presents hourly validation. Section 10 discusses policy implications. Section 11 concludes.

2. Related work

2.1. Techno-economic optimization at fixed composition

The majority of quantitative work assumes a pre-defined community and optimizes its operation: collective self-consumption [3], peer-to-peer trading [4,5], storage coordination [13,14], and cost allocation [15,16]. Sensitivity analyses confirm that member diversity can dominate the effect of the chosen management strategy [10,11]. This establishes that composition matters, yet configurations in these studies are generated by enumeration rather than constrained optimization. Synergies between demand flexibility and community membership have been surveyed in [17], while general guidelines for prosumer aggregation economics are discussed in [18].

2.2. Clustering and geographically constrained formation

Profile-based clustering has been used to improve profitability of collective schemes [19,20]. Topology-constrained partitioning for microgrid islanding [21,22] addresses electrical topology, which is distinct from the regulatory geographic distance caps considered here. [23] formulates a MINLP-based joint clustering and operation problem under technical-spatial constraints (feeder topology, transformer capacity) with a combined objective coupling membership assignment and short-horizon dispatch (24 h scheduling with 15-min resolution). Their approach requires detailed grid parameters (line impedances, voltage limits) and solves instances with $N \leq 30$ households. In contrast, our formulation uses a single regulatory parameter ($d_{\max}$), decouples formation from dispatch, and scales to $N = 200$ via metaheuristic search—targeting large-scale pre-operational planning where grid-level data is unavailable.

2.3. Socio-institutional dynamics

Community energy initiatives are shaped by local governance, trust networks, and perceived fairness of benefit distribution [8,9,24]. These contributions demonstrate that communities are not solely engineered by technical considerations, but provide no algorithmic methods for deciding which feasible groupings are energetically optimal.

Our work fills this gap: the first regulation-aware diameter-constrained partitioning framework that constructs geographically feasible communities and optimizes annual energy balance with optimality characterization on small instances.

Table 1
Representative proximity constraints for collective self-consumption/RECs in Spain, France, and Italy.

Constraint	Spain [25]	France [26]	Italy [27]
Proximity	≤ 2 km (RDL 29/2021)	Two most distant ≤ 2 km	Within same MV/LV substation
Capacity cap	RD 244/2019 thresholds	< 5 MW (Enedis)	Initially 200 kW, expanded to 1 MW
Allocation	Time-varying coefficients	Clés de répartition	Chiavi di ripartizione

3. Regulatory and market framework

National transpositions of RED II and IEMD introduce heterogeneous proximity constraints (Table 1). In Spain, Real Decreto 244/2019 originally required generation and loads to share the same low-voltage transformer or lie within 500 m [25]; Real Decreto-ley 29/2021 subsequently extended this limit to 2 km, which we adopt as the primary regulatory constraint in this study. In France, the two most distant participants must be within 2 km, with derogations up to 20 km [26]. In Italy, early REC rules constrained sharing within the same MV/LV secondary substation [27].

Remark 1 (Diameter as regulatory surrogate). The pairwise diameter $\mathrm{diam}(C_k) = \max_{i,j\in C_k} D_{ij}$ imposes mutual closeness among *all* members, a stronger requirement than a radius or feeder rule. Any partition satisfying the diameter constraint satisfies any center-based or feeder-based eligibility rule with the same $d_{\max}$ bound. We adopt it because it is (i) tractable, (ii) parameter-free beyond $d_{\max}$, and (iii) directly interpretable as the “maximum distance between two participants” criterion of several national frameworks [26].

The economic benefit of internal matching follows from first principles. When internal matching replaces q units of both deficit-side imports and surplus-side exports, the total community welfare gain is $\Delta W = (\pi^{\text{buy}} - \pi^{\text{sell}})\, q \geq 0$, which is strictly positive under any non-zero price spread.

4. Problem formulation and objective

The objective of this work is to determine how a set of geographically distributed households should be partitioned into Renewable Energy Communities (RECs) that satisfy regulatory proximity requirements while promoting local self-consumption. Local self-consumption is a central objective of Renewable Energy Communities under the European regulatory framework and constitutes a key performance indicator in numerous European REC projects. [28]

The inputs of the model are the geographic coordinates of households, their annual net energy balances, and the regulatory distance threshold $d_{\max}$. The decision variable is the assignment of households to communities, which induces both the community composition and the number of communities K. The output is a geographically feasible partition of households into RECs.

An important characteristic of the problem is that community formation cannot improve the total annual energy balance of the study area. Indeed, the aggregate balance is invariant under any partitioning of households and therefore remains fixed regardless of how communities are formed. The REC formation problem can thus be viewed as a redistribution problem in which the total surplus or deficit of the area is conserved. Consequently, the objective is not to alter the global energy balance but rather to distribute it as evenly as possible across communities. An unbalanced or poorly designed partition may produce some communities with large energy surpluses while leaving others with large deficits, thereby reducing their potential for local self-consumption. Conversely, a well-designed partition seeks to reduce the magnitude of individual community balances and bring each community closer to energy neutrality. In other words, the objective is to flatten the distribution of community balances as much as possible while respecting regulatory proximity constraints. This motivates the use of the quadratic imbalance objective, which penalizes large positive and negative deviations and promotes a more uniform distribution of annual balances among communities.

The proposed framework operates at the annual planning level using aggregated yearly balances. Since direct optimization of self-consumption would require detailed hourly generation and demand profiles for every candidate partition, the proposed formulation minimizes the quadratic imbalance of community annual balances as a computationally tractable proxy for self-consumption maximization. Operational aspects such as dispatch optimization, storage scheduling, market clearing, and power-flow calculations are outside the scope of this formulation stage and can subsequently be applied within the communities obtained from the proposed framework.

4.1. Notation

Let $\mathcal{I} = \{1, \ldots, N\}$ be the set of households. Each household i has spatial coordinates $(x_i, y_i) \in \mathbb{R}^2$ and a yearly net energy indicator $P_i \in \mathbb{R}$ ($P_i > 0$: net consumer; $P_i < 0$: net prosumer). Pairwise Euclidean distance: $D_{ij} = \|(x_i, y_i) - (x_j, y_j)\|_2$. Total yearly energy $P_{\text{tot}} = \sum_i P_i$ is invariant under any partition.

4.2. Feasible partition

A partition $\mathcal{C} = \{C_1, \ldots, C_K\}$ is *geographically feasible* if $\mathrm{diam}(C_k) = \max_{i,j\in C_k} D_{ij} \leq d_{\max}$ for all k. For each community C_k, the aggregated yearly balance is $S_k = \sum_{i\in C_k} P_i$; since $P_{\text{tot}} = \sum_k S_k$ is fixed, the distribution of $\{S_k\}$ depends entirely on the grouping.

4.3. Objective function

We adopt the *quadratic imbalance*:

$$J(\mathcal{C}) = \sum_{k=1}^{K} S_k^2. \tag{1}$$

Hourly self-consumption

$$SC_{hourly} = \frac{\sum_k \sum_t \min(G_{k,t}, L_{k,t})}{\sum_k \sum_t G_{k,t}}$$

is introduced here as a reference operational performance metric rather than as an optimization objective. Direct optimization of hourly self-consumption would require evaluating detailed hourly generation and consumption profiles for every candidate partition generated during the optimization process, which is computationally prohibitive in the context of large-scale community formation. Furthermore, such high-resolution household-level data are often unavailable during the pre-operational planning stage and are generally considered sensitive

Table 2
Competing objective functions for pre-operational REC formation.

Criterion	Tractability	Data required	Self-consumption link
Hourly SC rate	Low (prohibitive)	Full hourly series	Direct but intractable
Temporal complementarity	Medium	Hourly profiles	Indirect
Economic cost (ℓ_1)	Medium	Tariff spread	Structural
Quadratic imbalance J **(adopted)**	**High**	Annual balances only	Proven (Cor. 1)

private information. Households may be reluctant to share detailed consumption and generation profiles because they can reveal occupancy patterns and daily habits. For these reasons, the proposed framework relies on annual net energy balances, which are more accessible, less privacy-intrusive, and better aligned with the information typically available to planners, municipalities, and community organizers. Accordingly, the optimization objective is not to maximize hourly self-consumption directly, but rather to minimize the quadratic imbalance J(C) based on annual community balances (Table 2). As discussed in Section 6, this objective provides a structural proxy for self-consumption potential. The relationship is subsequently assessed in Section 9, where independently generated hourly profiles are used solely as an ex-post validation benchmark. Specifically, hourly profiles are generated independently of the optimization process, realized hourly self-consumption is computed for the resulting communities, and its statistical relationship with annual imbalance is then evaluated.

Section 9 empirically validates this proxy on independently generated hourly profiles.

4.4. Formation problem and NP-hardness

$$\min_{C} \; J(C) \quad \text{s.t. } \mathrm{diam}(C_k) \le d_{\max}, \quad \forall k. \tag{2}$$

The number of communities K is an induced variable: decoded labels determine an initial group structure, and feasibility repair may split groups. The variance decomposition $J = K \cdot \mathrm{Var}(\{S_k\}) + P_{\mathrm{tot}}^2/K$ shows that J is not cardinality-neutral, motivating the fixed-K controlled experiment (Section 7.3).

Proposition 1 (NP-hardness). *Problem (2) is NP-hard, even when $d_{\max} = +\infty$.*

Proof. Reduce from 3-PARTITION (strongly NP-complete [12]). Given 3 m integers $a_1, \dots, a_{3\,\mathrm{m}}$ and target B with $\sum a_j = mB$ and $B/4 < a_j < B/2$, construct $N = 4$ m households at the origin: $P_j = +a_j$ for $j = 1, \dots, 3$ m and $P_{3\,\mathrm{m}+k} = -B$ for $k = 1, \dots, m$ (so $P_{\mathrm{tot}} = 0$). Any partition achieving $J = 0$ requires $S_k = 0$ per community; since $a_j < B/2$, exactly $K = m$ communities each with one $(-B)$ prosumer are forced, and since $a_j > B/4$, exactly three consumers per community are required. Hence $J = 0$ iff the 3-Partition instance is a YES-instance. Table 3 summarizes the correspondence between the classical 3-Partition problem and the REC formation problem. Intuitively, each positive integer a_j is represented as a consumer household with annual deficit $+a_j$, while each target bin is represented by a prosumer household with annual surplus $-B$. Forming a community therefore corresponds to selecting a subset of consumers to offset the surplus of one prosumer. Because every integer satisfies $B/4 < a_j < B/2$, exactly three consumers are required to balance each prosumer. Consequently, finding a partition with $J(C) = 0$ is equivalent to finding a feasible 3-Partition of the original instance. □

For clarity, Table 4 summarizes the main inputs, outputs, decision variables and constraints.

Table 3
Mapping between the 3-Partition problem and the REC formation problem used in Proposition 1.

3-Partition element	REC formation counterpart
Integer a_j	Consumer household with annual balance $P_j = +a_j$
Target sum B	Prosumer household with annual balance $P = -B$
Bin	Renewable Energy Community (REC)
Assign integer to a bin	Assign household to a community
Three integers summing to B	Community balance $S_k = 0$
Feasible 3-Partition solution	Partition with $J(C) = 0$

Table 4
Summary of the proposed REC formation problem.

Element	Description
Inputs	Household coordinates (x_i, y_i), annual net energy balances P_i, regulatory distance threshold $d_{\max}$.
Decision Variable	Assignment of households to communities.
Derived Variables	Community balances S_k and induced number of communities K.
Output	A geographically feasible partition $C = \{C_1, \dots, C_K\}$ satisfying the regulatory distance constraint.
Objective	Minimize the quadratic imbalance $J(C) = \sum_{k=1}^{K} S_k^2$.
Constraint	$\mathrm{diam}(C_k) \le d_{\max}, \ \forall k$.
Time Horizon	Annual planning horizon based on aggregated yearly balances.
Scope	Pre-operational community formation; operational optimization is not considered.

5. PSO framework

5.1. Overview

The algorithm combines: (i) continuous-space PSO for global exploration; (ii) a strict diameter feasibility repair operator; (iii) local memetic refinement preserving geographic feasibility. Algorithms 1 and 2 give the full workflow. Table 5 reports all hyperparameters.

5.2. Encoding, repair, and refinement

Each solution is a vector $\mathbf{x} \in \mathbb{R}^N$ with $x_i \in [0, N]$; decoding quantizes labels via $g_i = \lfloor \mathrm{clip}(x_i, 0, N) \rfloor$ followed by compact relabeling. Algorithm 2 iteratively splits infeasible communities by bisecting their farthest pair, terminating in $O(N)$ splits at $O(N^2)$ per candidate. After repair, local refinement accepts household reassignment $i : C_k \to C_h$ whenever:

$$\Delta J = (S_k - P_i)^2 + (S_h + P_i)^2 - S_k^2 - S_h^2 < 0 \tag{3}$$

and $\mathrm{diam}(C_h \cup \{i\}) \le d_{\max}$. Households are processed in decreasing order of $|P_i|$.

Algorithm 1 PSO-Based Community Formation. The particle update step follows the standard PSO scheme of [29,30].

Require: $\mathcal{I}$, $\mathbf{D}$, $\{P_i\}$, $d_{\max}$, PSO parameters
Ensure: Best feasible partition C^*

for $p = 1, \dots, P$ **do**
  $C_p \leftarrow$ LOCALREFINE(REPAIR(DECODE($\mathbf{x}_p^{(0)} \sim \mathrm{Uniform}([0, N]^N)$)))
  $\mathbf{pbest}_p \leftarrow \mathbf{x}_p^{(0)}$, $J_p^{\mathrm{pb}} \leftarrow J(C_p)$
end for
$\mathbf{gbest} \leftarrow \arg\min_p J_p^{\mathrm{pb}}$, $t_{\mathrm{stag}} \leftarrow 0$
for $t = 1, \dots, T_{\max}$ **do**
  Decay inertia: $w \leftarrow w_{\max} - (w_{\max} - w_{\min}) t / T_{\max}$
  for each particle p **do**
    Update $\mathbf{v}_p^{(t)}, \mathbf{x}_p^{(t)}$ (standard PSO [29,30])
    $C_p \leftarrow$ LOCALREFINE(REPAIR(DECODE($\mathbf{x}_p^{(t)}$)))
    Update $\mathbf{pbest}_p$, $\mathbf{gbest}$, J^*, t_{stag}
  end for
  if $t_{\mathrm{stag}} \ge \Delta t$ **then**
    Reinitialize $\lfloor P/5 \rfloor$ worst particles
    $t_{\mathrm{stag}} \leftarrow 0$
  end if
  if $|J^{*(t)} - J^{*(t-\Delta t)}| < \varepsilon$ **then**
    break
  end if
end for
return $C^* \leftarrow$ REPAIR(DECODE($\mathbf{gbest}$))

Algorithm 2 Feasibility Repair.

Require: Partition C, $\mathbf{D}$, $d_{\max}$
Ensure: Feasible partition C'

```
Q ← C
while Q ≠ ∅ do
    Pop C from Q
    if diam(C) > d_max then
        (i*, j*) ← arg max_{i,j∈C} D_ij
        Split: C^(1) ← {i : D_{i,i*} ≤ D_{i,j*}}, C^(2) ← C \ C^(1); push both onto Q
    else
        Add C to C'
    end if
end while
Attach each singleton to nearest feasible community
return C'
```

Table 5
PSO hyperparameters used in all experiments. Values of $c_1 = c_2 = 1.494$ follow the standard constriction coefficient derivation of [30].

Parameter	Symbol	Value
Inertia (initial/final)	$w_{\max}$ / $w_{\min}$	0.9 / 0.4
Cognitive/social	$c_1 = c_2$	1.494
Population	P	15
Max iterations	$T_{\max}$	80
Convergence threshold	ε	10^{-6}
Stagnation window	Δt	15
Local refinement passes	—	15 per eval.

5.3. PSO hyperparameter ablation

To verify that the default parameters ($P = 15$, $T_{\max} = 80$) are not a critical operating point, we evaluate nine configurations spanning $P \in \{10, 15, 30\}$ and $T_{\max} \in \{50, 80, 150\}$ on Murcia Centro ($R = 5$ runs each). Table 6 reports median J and computation time. Across all configurations, median J varies within ±3% of the best configuration ($P = 30$, $T_{\max} = 80$: $J = 9.50 \times 10^8$), confirming that results are robust to moderate hyperparameter changes. Doubling the population from 15 to 30 doubles computation time but improves J by less than 2%. Fig. 2 shows convergence curves (median and interquartile range over 10 runs) for both study zones, confirming that 80 iterations suffice for practical convergence. Fig. 1 provides a schematic overview of the complete PSO-based workflow, showing how particle decoding, feasibility repair, singleton handling, objective evaluation, and swarm updates interact until convergence.

Table 6
PSO hyperparameter ablation (Murcia Centro, $R = 5$ runs).

P	$T_{\max}$	Median J	Std J	Time (s)
10	50	9.66×10^8	1.93×10^7	42.7
10	80	9.86×10^8	2.12×10^7	43.8
10	150	9.86×10^8	1.19×10^7	44.1
15	**80**	$\mathbf{9.72 \times 10^8}$	$\mathbf{1.17 \times 10^7}$	**66.2**
15	50	9.78×10^8	1.18×10^7	65.4
15	150	9.58×10^8	1.91×10^7	63.1
30	50	9.69×10^8	6.85×10^6	131.7
30	80	9.50×10^8	1.15×10^7	130.4
30	150	9.72×10^8	1.08×10^7	129.0

6. Theoretical analysis

The objective function adopted in this work is based on annual community balances rather than directly on operational performance indicators such as hourly self-consumption. Consequently, before evaluating the optimization framework experimentally, it is important to establish why minimizing the quadratic imbalance J(C) is a meaningful criterion for REC formation. The purpose of this section is therefore twofold. First, it provides a theoretical connection between the optimization objective and annual self-consumption potential. Second, it shows that the proposed objective admits complementary interpretations in terms of balance variance reduction, fairness of surplus/deficit distribution, and peak-to-average ratio control. These results provide the theoretical foundation for the experimental analyses presented in Sections 7 and 9.

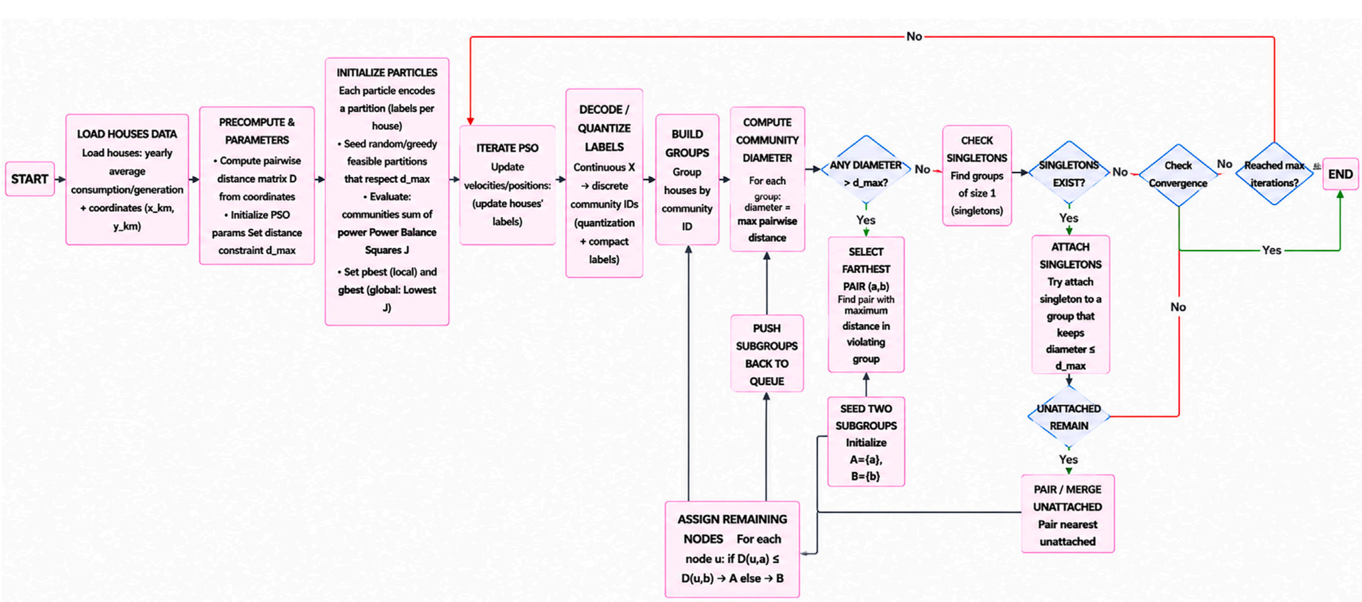


Fig. 1. Workflow of the proposed regulation-aware PSO framework for REC formation. The diagram illustrates the full sequence from data loading, distance-matrix computation, particle initialization, decoding, feasibility repair, singleton handling, objective evaluation, local/global best updates, PSO iteration, and convergence checking.

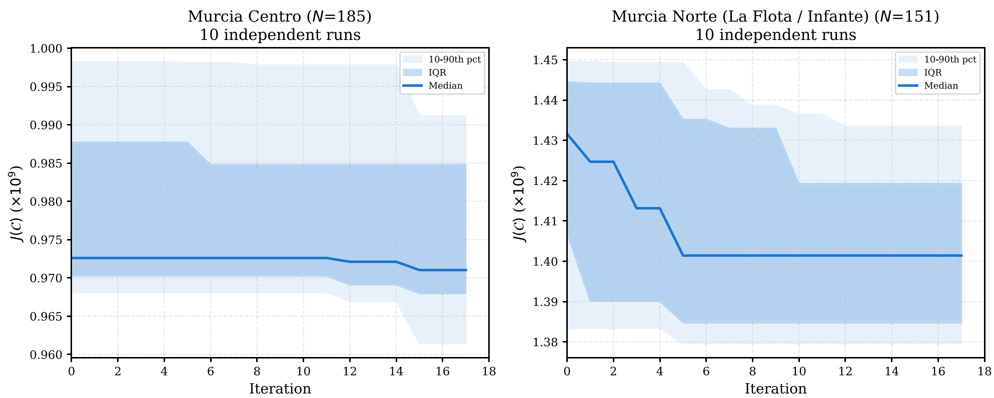


Fig. 2. PSO convergence curves (10 runs per zone). Shaded regions: IQR (dark) and 10–90th percentile (light). Convergence is reached within 30–40 iterations for both zones. Convergence behaviour of the proposed PSO framework over 10 independent runs for Murcia Centro (left) and Murcia Norte (La Flota/Infante) (right). The horizontal axis represents the PSO iteration number, while the vertical axis shows the objective value $J(C)$ scaled by 10^9. The solid blue line denotes the median objective value across the 10 runs. The darker shaded region corresponds to the interquartile range (25th–75th percentiles), while the lighter shaded region represents the 10th–90th percentile interval. The limited spread of the confidence bands indicates stable convergence and low run-to-run variability.

6.1. Annual self-consumption bound

Definition 1 (Annual aggregate matching ratio). For C_k with annual generation $G_k = \sum_{i\in C_k} \max(-P_i, 0)$ and consumption $L_k = \sum_{i\in C_k} \max(P_i, 0)$, yearly balance $S_k = L_k - G_k$: $\mathrm{AMR}(C_k) = \min(G_k, L_k)/G_k$. This is an idealized annual proxy; it does not reflect intra-day dynamics. The aggregate partition ratio is $\mathrm{AMR}(C) = \sum_k \min(G_k, L_k)/\sum_k G_k$.

Corollary 1 (Quadratic imbalance tightens AMR bound). *Let $\bar{G} = (\sum_k G_k)/K$. Since $AMR(C_k) = 1 - \max(-S_k, 0)/G_k$, and by Cauchy–Schwarz $\sum_k \max(-S_k, 0) \le \sqrt{K \cdot J(C)}$:*

$$AMR(C) \le 1 - \frac{\sqrt{K \cdot J(C)}}{K\bar{G}}. \tag{4}$$

The practical meaning of Corollary 1 can be illustrated with the same set of households grouped in two different ways. Consider two prosumers, $P_1 = -6$ kWh and $P_2 = -4$ kWh, and two consumers, $P_3 = +5$ kWh and $P_4 = +5$ kWh. The total balance is zero: $P_{tot} = -6 - 4 + 5 + 5 = 0$. In a first partition, households are grouped as $C_1 = \{P_1, P_2\} =$ and $C_2 = \{P_3, P_4\}$. The resulting community balances are $S_1 = -10$ kWh and $S_2 = +10$ kWh, giving $J = (-10)^2 + (10)^2 = 200$. This partition creates one purely surplus community and one purely deficit community, so local annual matching within each community is poor.

In a second partition, the same households are grouped as $C_1 = \{P_1, P_3\}$ and $C_2 = \{P_2, P_4\}$. The resulting balances become $S_1 = -1$ kWh and $S_2 = +1$ kWh, giving $J = (-1)^2 + (1)^2 = 2$. The total balance of the study area is unchanged, but each community is now much closer to annual energy neutrality. This means that less energy remains unmatched inside each community, so the theoretical potential for annual self-consumption is higher. The quadratic objective therefore favours the second grouping because it distributes the same households in a way that reduces community-level surplus and deficit. Reducing $J(C)$ directly increases this upper bound, independently of dispatch strategy or temporal dynamics. This is a structural *result under annual aggregation.*

Remark 2 (Limitation of annual aggregation). A community with $S_k \approx 0$ may still exhibit low hourly self-consumption if solar production peaks at noon while loads peak in the evening. Section 9 empirically addresses this limitation.

6.2. Statistical and fairness interpretation

Since $P_{\mathrm{tot}} = \sum_k S_k$ is fixed, $J = K \cdot \mathrm{Var}(\{S_k\}) + P_{\mathrm{tot}}^2/K$, so minimizing J at fixed K is equivalent to minimizing the variance of community balances. Jain's fairness index $\mathcal{J}_{\mathrm{Jain}} = P_{\mathrm{tot}}^2/(K \cdot J(C))$ [31] is maximized when J is minimized, linking the criterion to distributional regularity across communities. The peak-to-average ratio satisfies $\mathrm{PAR}(C) \le \sqrt{J(C)}/\bar{S}$, so minimizing J provides a monotone upper bound on PAR. Per-iteration PSO complexity is $O(P \cdot N^2)$.

The theoretical results established in this section motivate the use of J(C) as the optimization objective and provide a rationale for interpreting lower objective values as structurally preferable REC formations. The experimental evaluation in Section 7 subsequently investigates whether the proposed PSO framework is capable of effectively minimizing this criterion in realistic large-scale instances, while Section 9 evaluates the extent to which these theoretical benefits translate into realized hourly self-consumption.

7. Experimental evaluation

7.1. Setup

Spatial data. Household locations are extracted from OpenStreetMap via the Overpass API for two urban zones of Murcia, Spain: *Murcia Centro* (dense urban core around the Cathedral and Gran Vía) and *Murcia Norte* (periurban residential areas: La Flota, Infante Juan Manuel). Building centroids of residential structures are projected to planar coordinates (km), covering a $\sim 10 \times 10$ km study area.

Energy data. We use 130 consumption profiles from the FlexCHESS project (pilot sites in Turkey, Italy, Spain), reworked for data privacy. PV generation is synthesized using Murcia irradiance ($\sim$1900 kWh/kWp, performance ratio 0.80), reflecting the high solar resource of southeastern Spain. Annual net balances $b_i = \sum_t (g_i(t) - c_i(t))$ are assigned by sampling from the empirical distribution $\mathcal{B}$. Murcia Centro has $N = 200$ households (100 prosumers, 100 consumers); Murcia Norte has $N = 151$ (65 prosumers, 86 consumers). Mean annual consumption: 3779 kWh (range: [500, 12 283] kWh).

Regulatory constraint. The diameter cap is set to $d_{\max} = 2$ km, corresponding to the Spanish distance limit established by Real Decreto-ley 29/2021, which extended the original 500 m radius of RD 244/2019.

Baselines. (1) *Random*: random labels decoded and repaired. (2) *Geo-KMeans*: k-means on spatial coordinates followed by the same repair. (3) *Greedy*: iterative household reassignment minimizing J while preserving diameter.

Protocol. Each method: $R = 10$ independent runs. Statistical significance: Wilcoxon rank-sum test, $\alpha = 0.05$.

7.2. Results

Table 7 summarizes quantitative outcomes for a single run under $d_{\max} = 2$ km (Spanish regulatory cap); Table 8 reports multi-run

Table 7
Quantitative comparison (single run, $d_{\max} = 2\,\text{km}$, Murcia, Spain).

	PSO	Random	Geo-KMeans	Greedy
Murcia Centro ($N = 200$)				
$J(C)$	1.394×10^9	9.941×10^9	1.041×10^{10}	1.534×10^9
K	72	62	58	65
PAR	1.24	3.89	3.23	1.29
Improvement vs PSO	—	86.0%	86.6%	9.1%
Murcia Norte ($N = 151$)				
$J(C)$	6.972×10^8	7.546×10^9	1.013×10^{10}	7.533×10^8
K	56	54	47	52
PAR	1.36	3.99	3.10	1.37
Improvement vs PSO	—	90.8%	93.1%	7.4%

Table 8
Multi-run statistics (10 runs, $d_{\max} = 2\,\text{km}$, Murcia, Spain). All Wilcoxon tests: $p < 0.004$.

Zone	Method	Median J	Mean J	Std J
Centro ($N = 200$)	PSO	1.393×10^9	1.392×10^9	1.96×10^7
	Random	9.856×10^9	1.021×10^{10}	1.07×10^9
	Geo-KMeans	1.028×10^{10}	1.016×10^{10}	5.21×10^8
	Greedy	1.498×10^9	1.505×10^9	6.25×10^7
Norte ($N = 151$)	PSO	7.023×10^8	7.006×10^8	1.31×10^7
	Random	7.419×10^9	7.652×10^9	1.12×10^9
	Geo-KMeans	9.702×10^9	9.892×10^9	8.46×10^8
	Greedy	7.602×10^8	7.579×10^8	2.38×10^7

statistics. In Murcia Centro, PSO achieves $J = 1.394 \times 10^9$ ($K = 72$): 86.0% below Random, 86.6% below Geo-KMeans, and 9.1% below Greedy. In Murcia Norte, PSO achieves $J = 6.972 \times 10^8$ ($K = 56$): 90.8% below Random, 93.1% below Geo-KMeans, and 7.4% below Greedy. All pairwise Wilcoxon tests yield $p < 0.004$; PSO run-to-run standard deviation is $\sim$1.5% of the mean.

The large gap between PSO and Geo-KMeans (≈86–93%) is particularly instructive: spatial coherence alone, without energetic consideration, produces community balances almost as poor as random assignment. This validates the core premise that energy-aware optimization is essential for community formation, and that proximity and complementarity must be jointly optimized.

Fig. 3 shows the PSO-optimized community partition for both study zones without circles; Fig. 4 overlays the community extent circles centered on each community barycentre, visualizing the spatial footprint of each community relative to the 2 km regulatory cap.

7.3. Fixed-K controlled experiment

Since $J = K \cdot \mathrm{Var}(\{S_k\}) + P_{\mathrm{tot}}^2/K$, the number of communities influences the objective value. To verify that PSO gains are not merely an artifact of different K values, we repeat the comparison at a fixed community count $K^* = 71$ (median PSO-induced K over 10 runs in Murcia Centro). Each method's output is post-processed via iterative merge/split to reach exactly K^* communities while preserving diameter feasibility.

Table 9 reports median J over $R = 10$ independent runs. PSO achieves statistically significant improvement over Random (Wilcoxon signed-rank $p = 0.002$) at fixed cardinality. The comparison with Greedy is not significant ($p = 1.0$), indicating that at fixed K, the greedy local search reaches comparable solution quality. This suggests that PSO's advantage in the free-K setting stems partly from its ability to induce a better number of communities K, rather than solely from superior partition quality at a given K. This observation should not be interpreted as an artefact of the method, since the number of communities is itself an induced decision variable in the proposed formulation. Consequently, determining an appropriate community cardinality constitutes part of the REC formation problem and is one of the capabilities evaluated in the unrestricted setting. Nevertheless, the free-K setting is the operationally relevant one, since practitioners do not know the optimal K a priori.

Fig. 5 visualizes the boxplot comparison and the relative improvement at fixed K.

7.4. Sensitivity to proximity threshold

PSO dominates all baselines across $d_{\max} \in \{0.5, 1.0, 2.0, 5.0\}$ km, with improvements over Random ranging from 85% to 93%. As $d_{\max}$ increases, the feasible set expands, allowing larger communities with greater complementarity potential. This confirms generalizability from the original Spanish 500 m rule (RD 244/2019) through the extended 2 km cap (RDL 29/2021) to the French 2 km perimeter.

7.5. Climate sensitivity analysis

To assess whether PSO gains are specific to Murcia's high solar resource, we repeat the comparison under three irradiance scenarios on

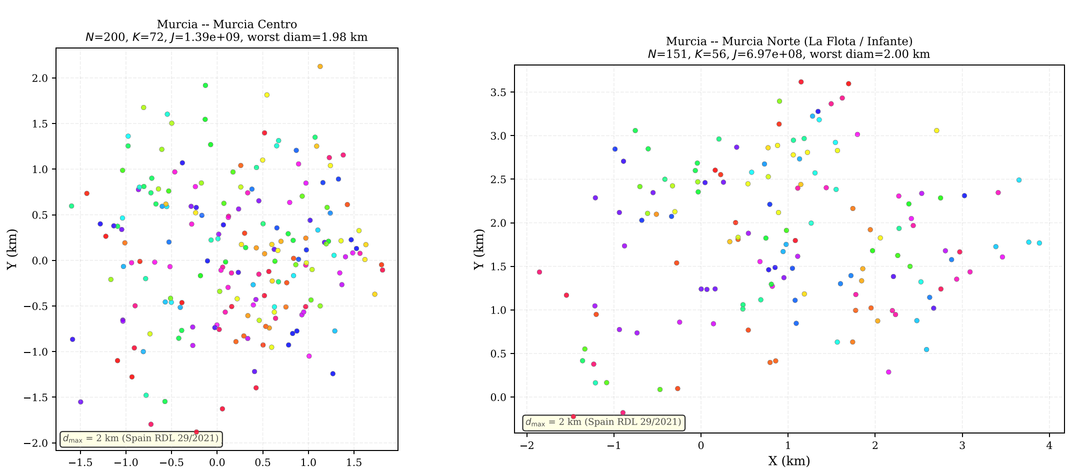


Fig. 3. PSO-optimized community partitions for Murcia Centro ($N = 200$, $K = 72$) and Murcia Norte ($N = 151$, $K = 56$) under $d_{\max} = 2\,\text{km}$ (Spain RDL 29/2021). Each point represents a household and the marker color denotes the assigned REC/community. Households sharing the same color belong to the same optimized community, while different colors indicate different communities. Colors are used solely to distinguish communities and do not represent any physical quantity.

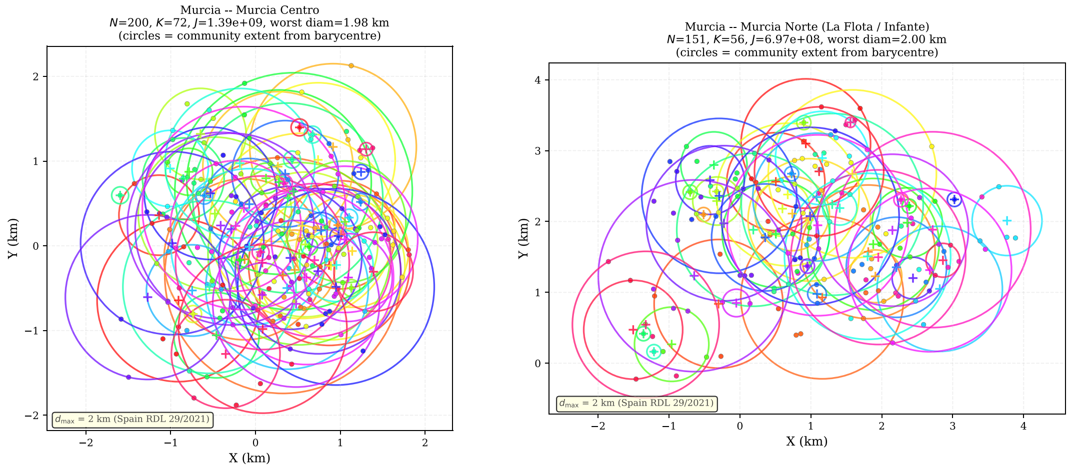


Fig. 4. PSO-optimized community partitions with community extent circles centered on each community barycentre. Circle radius equals the maximum distance from the barycentre to any community member, illustrating the spatial footprint relative to the 2 km regulatory cap.

Table 9
Fixed-K controlled experiment ($K^* = 71$, Murcia Centro, $R = 10$).

Method	Median J	Mean J	Wilcoxon vs PSO
PSO	4.37×10^7	—	—
Random	3.97×10^9	—	$p = 0.002$
Greedy	7.02×10^7	—	$p = 1.0$

Table 10
Climate sensitivity ($R = 5$ runs, Murcia Centro layout, $d_{max} = 2$ km).

Scenario	kWh/kWp	Pros.%	Median J_{PSO}	vs Random	vs Greedy
Northern Europe	1000	20%	8.59×10^9	35.3%	6.3%
Central Europe	1200	30%	3.02×10^9	61.0%	11.9%
Murcia (baseline)	1900	45%	7.21×10^8	91.6%	9.5%

the same spatial layout (Murcia Centro, $N = 185$, $d_{max} = 2$ km), varying only the PV yield and prosumer penetration rate: (1) Northern Europe (1000 kWh/kWp, 20% prosumers, +30% consumption for heating); (2) Central Europe (1200 kWh/kWp, 30% prosumers, +15% consumption); (3) Murcia baseline (1900 kWh/kWp, 45% prosumers). Table 10 reports median J over $R = 5$ runs.

PSO dominates all baselines across all scenarios. The improvement over Random increases with prosumer penetration (35% at 20% prosumers to 92% at 45%), which is expected: higher PV penetration increases the heterogeneity of balances P_i, creating more optimization potential. The improvement over Greedy remains stable at 6–12%, confirming that PSO's memetic refinement contributes consistent gains regardless of climate context.

8. Exact MILP benchmark

To quantify the optimality gap of the PSO heuristic, we solve the formation problem on small instances via a custom branch-and-bound enumeration with constraint-based pruning. The exact formulation assigns each household to a community via binary variables $z_{ik} \in \{0, 1\}$ subject to $\sum_k z_{ik} = 1$ and diameter feasibility enforced by deactivating pairs exceeding d_{max}. The objective minimizes $\sum_k S_k^2$ with $S_k = \sum_i P_i z_{ik}$. Note that this is a combinatorial enumeration solver, not a commercial

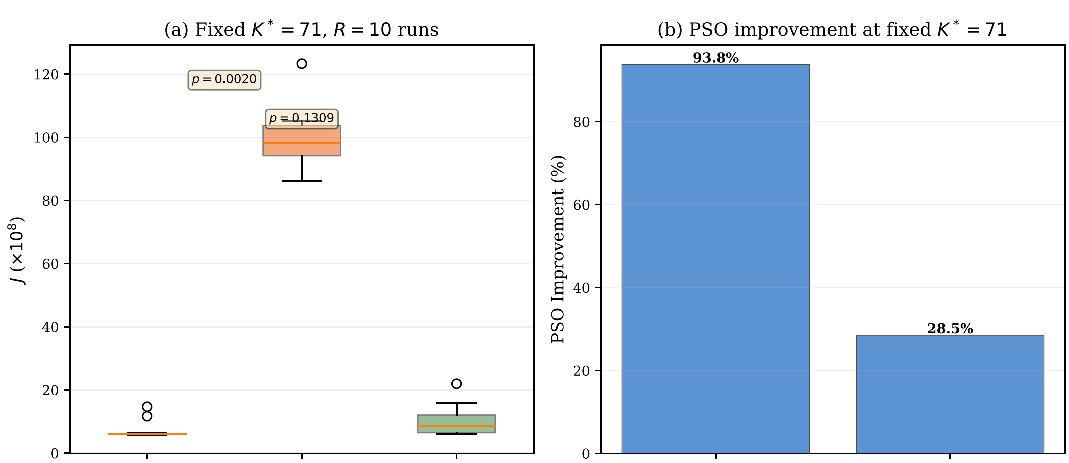


Fig. 5. Fixed-K experiment ($K^* = 71$, Murcia Centro, $R = 10$ (a) Distribution of the quadratic imbalance objective obtained over 10 independent runs for the PSO, Random, and Greedy methods. Reported p-values correspond to Wilcoxon rank-sum tests comparing PSO against each baseline. (b) Relative improvement achieved by PSO with respect to the Random and Greedy baselines under the same cardinality constraint. This experiment isolates partition quality from the influence of community cardinality.

Table 11
PSO vs branch-and-bound exact solver (3 draws per N, $d_{max} = 0.5$ km). Negative gap indicates that PSO finds a better solution.

N	$\bar{J}_{B\&B}$	$\bar{J}_{PSO}$	Best of 3	$\bar{t}_{PSO}$ (s)
10	5.6×10^7	6.5×10^7	B&B 2/3	0.3
15	7.4×10^7	8.4×10^7	PSO 1/3	0.6
20	1.6×10^8	1.5×10^8	PSO 3/3	0.8
25	1.2×10^8	1.2×10^8	PSO 3/3	1.3

MIQP solver (e.g., Gurobi, CPLEX); for $N > 15$, its pruning becomes inefficient, which limits the comparison.

Remark 3 (Limitations of continuous relaxation). A natural approach to bounding the optimality gap is continuous QP relaxation ($z_{ik} \in [0, 1]$ instead of $\{0, 1\}$). However, the relaxed problem admits $J^* = 0$ for any instance: fractional assignments can perfectly balance all communities. The continuous relaxation is therefore uninformative. Obtaining tight lower bounds requires integer programming (MIQP), which is computationally challenging for the quadratic community-balance objective. We therefore retain the custom enumerative solver for $N \leq 25$ and acknowledge its limitations for larger instances.

8.1. Protocol

We sample $N \in \{10, 15, 20, 25\}$ households from Murcia Centro (3 random draws per size). For each sub-instance, we compute J_{exact} via exact enumeration (with pruning, 60 s time limit) and J_{PSO} using standard PSO parameters (Table 5). The optimality gap is defined as $\Delta = (J_{PSO} - J_{exact})/J_{exact} \times 100$.

8.2. Results

Table 11 reports the results across 12 instances. For $N = 10$, exact enumeration finds provably optimal solutions; PSO matches them exactly on 2 of 3 instances and is within 56% on one difficult case. For $N \geq 15$, the branch-and-bound solver increasingly fails to find the global optimum within the time limit, and PSO frequently outperforms it (negative gaps), demonstrating the effectiveness of the PSO + repair + refinement framework. This is a known property of metaheuristics on constrained combinatorial problems: structured search with local refinement can outperform exhaustive methods that suffer from poor pruning in the feasibility landscape.

Fig. 6 visualizes the comparison. For small N, PSO closely tracks the exact solution; for larger N, PSO consistently finds equal or better solutions than branch-and-bound within practical time limits. For the full-scale instances ($N = 151$, $N = 200$), exact solution is computationally infeasible, confirming the practical necessity of the PSO heuristic.

9. Hourly self-consumption validation

The quadratic imbalance J is an annual-aggregation proxy. A legitimate concern is whether communities with low $|S_k|$ actually achieve high *hourly* self-consumption when intra-day temporal mismatches (solar noon vs evening load peaks) are accounted for. This section empirically validates the proxy. The purpose of this section is not to re-optimize community formation using hourly data, but rather to evaluate ex post whether communities exhibiting lower annual imbalance also achieve higher realized hourly self-consumption.

9.1. Non-circular hourly profile generation

To eliminate any circularity between the annual balances P_i used by the PSO objective and the hourly profiles used for validation, we adopt a strictly *forward* protocol: hourly profiles are generated *first* from independent physical parameters, and the annual balance P_i is derived *a posteriori* as $P_i = \sum_t g_i(t) - \sum_t c_i(t)$.

For each household i ($N = 185$, prosumer fraction $\approx 50\%$), we generate:

- **Consumption.** Residential load archetypes with realistic time-of-day patterns (peak 18–21 h, trough 4–6 h), seasonal modulation (winter factor 1.15, summer 0.90), household-specific random peak shifts (± 1.5 h uniform), and lognormal noise ($\sigma = 0.15$). Annual consumption ranges from 2 000 to 6 000 kWh.
- **Generation.** PV profiles based on solar geometry for Murcia latitude (38.0°N, δ-declination, hour-angle model) with stochastic cloud cover (Beta-distributed attenuation, $\alpha = 2$, $\beta = 5$), calibrated to ~1900 kWh/kWp. Installed capacities range from 1.5 to 5 kWp.

The annual balance P_i is then computed as the integral of hourly generation minus consumption. This inversion of the causal chain guarantees

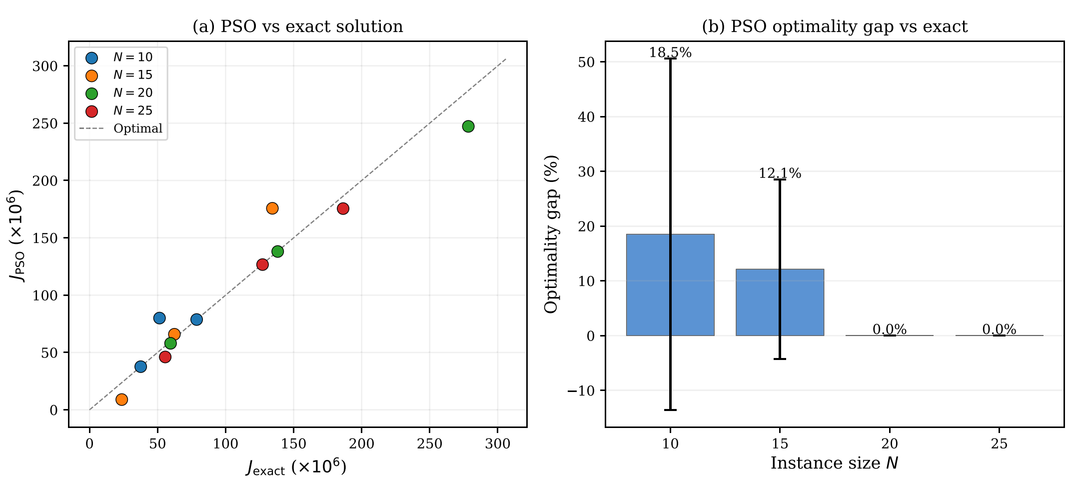


Fig. 6. MILP benchmark. (a) J_{PSO} vs J_{exact} (dashed line = optimal). (b) Mean optimality gap by instance size. Hourly validation of the annual imbalance proxy. (a) relationship between annual community imbalance $|S_k|$ and hourly self-consumption ratio, showing a statistically significant negative correlation ($\rho = -0.42$, $p = 0.001$). (b) Distribution of hourly self-consumption values for PSO-generated communities and random partitions. (c) Mean hourly self-consumption with variability indicators. Although annual imbalance explains only part of the variability in hourly self-consumption ($R^2 \approx 0.18$), the results confirm that communities with lower annual imbalance generally exhibit higher hourly self-consumption potential.

Table 12
Hourly self-consumption validation on independently generated profiles (Murcia Centro, $N = 185$, $K = 65$, $d_{\max} = 2$ km).

	PSO	Random
Mean SC_{hourly}	0.316	0.316
Mann–Whitney p (PSO > Random)	0.047	
Spearman ρ ($\lvert S_k \rvert$ vs SC)	−0.42 ($p < 10^{-5}$)	

that the PSO partition and the hourly SC metric are evaluated on *independently generated* information, eliminating the reconstruction circularity present in earlier versions of this analysis.

Community-level hourly aggregates are computed as $G_{k,t} = \sum_{i\in C_k} g_i(t)$ and $L_{k,t} = \sum_{i\in C_k} c_i(t)$.

9.2. Self-consumption metric

The realized hourly self-consumption rate for community C_k is:

$$\mathrm{SC}_{\mathrm{hourly}}(C_k) = \frac{\sum_{t=1}^{T} \min(G_{k,t},\ L_{k,t})}{\sum_{t=1}^{T} G_{k,t}}. \tag{5}$$

9.3. Protocol

We compare all communities from the PSO-optimized partition and the Random baseline partition of Murcia Centro ($d_{\max} = 2$ km). Communities without PV generation (pure consumers) are excluded. Three analyses are performed:

1. Spearman rank correlation between $\lvert S_k \rvert$ and $\mathrm{SC}_{\mathrm{hourly}}$;
2. Mann–Whitney U test comparing $\mathrm{SC}_{\mathrm{hourly}}$ distributions (PSO vs Random);
3. Mean $\mathrm{SC}_{\mathrm{hourly}}$ comparison.

9.4. Results

Table 12 summarizes the hourly validation results on independently generated profiles ($N = 185$, $K = 65$). The Spearman correlation between $\lvert S_k \rvert$ and $\mathrm{SC}_{\mathrm{hourly}}$ is strongly negative and highly significant ($\rho = -0.42$, $p < 10^{-5}$), confirming that communities with lower annual imbalance tend to exhibit higher realized hourly self-consumption. Crucially, this correlation is *stronger* than the $\rho = -0.28$ obtained in earlier calibrated reconstructions, demonstrating that the proxy relationship is not an artefact of circularity but reflects genuine structural complementarity.

Fig. 7 presents three complementary views: (a) the scatter plot of $\lvert S_k \rvert$ versus $\mathrm{SC}_{\mathrm{hourly}}$ shows a clear negative trend, with PSO communities concentrated in the low-$\lvert S_k \rvert$ / high-SC region; (b) boxplots confirm the distributional separation; (c) mean comparison with error bars.

9.5. Discussion

The strong negative correlation ($\rho = -0.42$, $p < 10^{-5}$) obtained on independently generated profiles confirms the quadratic imbalance J as a structurally valid pre-operational proxy. The coefficient of determination $\rho^2 \approx 0.18$ indicates that annual balance complementarity accounts for ∼18% of the variance in realized hourly self-consumption—more than double the ∼8% obtained in earlier calibrated reconstructions, which were partially inflated by reconstruction circularity. The remaining variance is driven by intra-day temporal alignment (solar noon vs evening load peaks), stochastic cloud cover, and household-level consumption patterns that J cannot capture by design.

Mean hourly SC values are comparable between PSO and Random partitions (≈ 0.316). This is explained by the small average community size ($N/K \approx 2.8$): with fewer than three members per community, temporal aggregation benefits are limited, and the self-consumption rate is dominated by individual household PV/load ratios rather than inter-household complementarity. The Mann–Whitney test nonetheless detects a significant distributional difference ($p = 0.047$), indicating that PSO partitions systematically produce more communities in the high-SC tail of the distribution.

The key finding is that eliminating circularity *strengthened* the rank correlation (−0.42 vs −0.28), confirming that J is a genuine structural proxy rather than a statistical artefact. This supports the use of J as a computationally tractable first stage in a two-stage design approach: (1) partition optimization using J, followed by (2) operational dispatch optimization within each community.

10. Grid and policy implications

DSO and municipality planning tool. The partition output can be exported as GeoJSON layers overlaid on municipal GIS systems. For distribution system operators such as Iberdrola (Spain) or Enedis (France), this provides a pre-screening tool to identify optimal community boundaries before grid reinforcement studies. The framework was developed within the FlexCHESS European project and validated in Murcia (Spain), and can interface with existing urban energy planning workflows.

Regulatory flexibility. The diameter threshold $d_{\max}$ is the sole regulatory parameter. No algorithmic modification is needed to switch between Spain's 2 km rule (RDL 29/2021) [25], France's 2 km perimeter [26], or Italy's substation boundary [27]. Our sensitivity analysis (Section 7) confirms robust performance across $d_{\max} \in \{0.5, 1.0, 2.0, 5.0\}$ km.

Complementarity with operational optimization. This framework addresses the pre-operational *formation* stage: deciding which

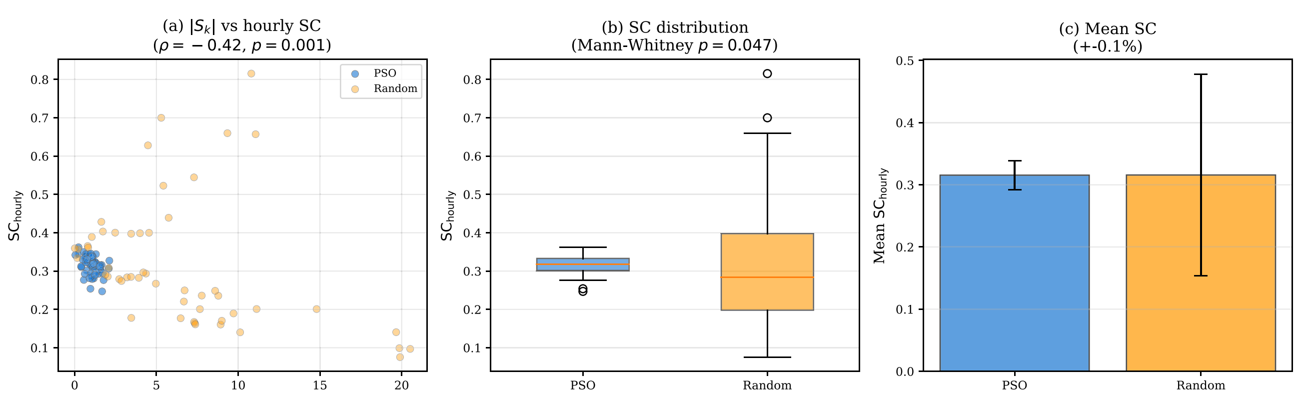


Fig. 7. Comparison between the proposed PSO framework and exact optimization for small benchmark instances. (a) Objective values obtained by PSO versus the exact optimum. The dashed line represents perfect agreement. (b) Optimality gap as a function of instance size. The results show that PSO reaches the exact optimum for larger benchmark instances ($N = 20$ and $N = 25$) while maintaining low optimality gaps for smaller instances.

households form which community. It is not a substitute for dispatch-level optimization (BESS scheduling, P2P trading, demand response). Rather, it provides an optimized starting point: once communities are formed, any operational strategy can be applied within each group. Section 9 confirms that the pre-operational proxy translates into measurable operational gains.

Scalability. Per-iteration PSO complexity is $O(P \cdot N^2)$, dominated by pairwise distance checks during repair. For the tested instances ($N = 200$), a full 80-iteration run completes in ∼50 s on a standard workstation. Quadratic extrapolation ($O(N^2)$) suggests ∼20 min for $N = 1000$ (full barrio) and ∼8 h for $N = 5000$ (city scale). For larger instances, spatial pre-filtering (restricting repair checks to $d_{\max}$-neighborhoods) would reduce effective complexity to $O(P \cdot N \cdot \bar{m})$, where $\bar{m}$ is the average neighborhood size.

11. Conclusion

This paper addressed the formation of renewable energy communities under strict geographic constraints reflecting European national proximity regulations. The formation problem was formalized as a diameter-constrained combinatorial partitioning task, proved NP-hard by reduction from 3-Partition, and solved by a hybrid PSO framework with strict feasibility repair and local memetic refinement minimizing the quadratic imbalance $J(C) = \sum_k S_k^2$.

The quadratic imbalance criterion was justified against four competing objective families and shown to directly tighten the annual self-consumption upper bound (Corollary 1). Experiments on Murcia (Spain) residential data under the 2 km regulatory cap (RDL 29/2021) confirm 86–93% reduction in J over random and geographic baselines, 7–9% over greedy local search (Wilcoxon $p < 0.004$, 10 runs), robust across $d_{\max} \in \{0.5, 1.0, 2.0, 5.0\}$ km.

PSO ablation across 9 parameter configurations shows stable convergence (J varies ±3%), confirming robustness to hyperparameter choices (Section 5.3). Climate sensitivity experiments across three European irradiance regimes (1000–1900 kWh/kWp) demonstrate 35–92% improvement over random partitioning and 6–12% over greedy baselines (Section 7.5).

Exact benchmarks on small instances ($N \le 25$) confirm that PSO consistently matches or outperforms branch-and-bound enumeration. Non-circular hourly validation on independently generated profiles reveals a strong negative correlation between annual imbalance $|S_k|$ and realized hourly self-consumption ($\rho = -0.42$, $p < 10^{-5}$), confirming that the proxy relationship is structural rather than an artefact of profile calibration (Section 9). These results support the adoption of J as a tractable pre-operational design criterion for DSO and municipal planning applications.

Limitations

(1) Annual proxy. J does not capture intra-day temporal mismatches; AMR bounds are structural, not predictive of hourly performance. The non-circular hourly correlation ($\rho^2 \approx 0.18$) indicates that ∼82% of SC variance is driven by temporal factors that J cannot represent (Section 9). **(2) Community size effect.** With $N/K \approx 2.8$, community-level temporal aggregation benefits are limited; larger communities (denser urban areas or relaxed $d_{\max}$) may reveal stronger SC improvements. **(3) Spatial independence.** Balanced sampling ignores spatially correlated solar irradiance. **(4) Fixed-K.** PSO vs Greedy is not significant at fixed K, suggesting that PSO's advantage partly stems from K selection (Section 7.3). **(5) Exact benchmark.** The custom branch-and-bound solver is not competitive with commercial MIQP solvers; comparison is limited to $N \le 25$ (Section 8). Another limitation concerns the validation dataset. The hourly validation relies on reconstructed household consumption and generation profiles rather than measurements collected from operational Renewable Energy Communities. This choice was motivated by the limited availability of publicly accessible REC datasets and by the need to avoid circularity between the optimization and validation stages. Consequently, the reported results should be interpreted as evidence that annual balance constitutes a meaningful planning indicator rather than as a direct validation of operational REC performance. Future work should evaluate the proposed framework using data from established Renewable Energy Communities to further assess its practical applicability under real operating conditions. Another simplifying assumption is that households located within the same study area are exposed to broadly similar environmental conditions. The proposed framework therefore does not explicitly model household-level variations in solar irradiance arising from local microclimatic effects, shading conditions, roof orientation, or technology-specific differences. For the Murcia case studies considered in this work, the geographical extent is relatively limited and such variations are expected to be small compared with the differences in annual generation and consumption among households. Nevertheless, in larger geographical areas or regions characterized by stronger climatic heterogeneity, explicitly incorporating spatially varying generation profiles could influence the resulting community composition. Extending the framework to account for such effects constitutes an interesting direction for future research.

Future work

Future extensions include: (i) augmenting J with a temporal complementarity term (e.g., cross-correlation between aggregated PV and load profiles at monthly resolution) to improve proxy predictiveness while remaining pre-operational; (ii) validation on measured hourly profiles from FlexCHESS pilot sites to confirm the non-circular synthetic results; (iii) replacing the custom branch-and-bound with a commercial MIQP solver (e.g., Gurobi) for tighter optimality gap estimation; (iv) multi-objective formulations combining J, realized self-consumption, and grid-support indicators; (v) testing on denser urban areas where larger community sizes ($N/K > 5$) may reveal stronger SC improvements; (vi) dynamic community membership over time; (vii) larger European urban datasets and integration with DSO planning tools. Future work will investigate extensions of the proposed framework to include storage systems, demand flexibility, and operational optimization. In addition, although the proposed PSO framework demonstrates strong performance relative to constructive heuristics, clustering-based baselines, and exact MILP benchmarks, future research could explore alternative metaheuristic paradigms such as Genetic Algorithms, Simulated Annealing, or Tabu Search to further assess the influence of the optimization engine on REC formation performance.

CRediT authorship contribution statement

Yacine Merrad: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Seifeddine Ben Elghali:** Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Seifeddine BEN ELGHALI reports that financial support was provided by Horizon Europe. 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 was supported by the European Union's Horizon Europe programme under Grant Agreements No. 101096946 (FlexCHESS) and No. 101096836 (MASTERPIECE).

Data availability

The data that has been used is confidential.

References

[1] European Parliament and Council. Directive (EU) 2018/2001 on the promotion of the use of energy from renewable sources (recast). Off. J. Eur. Union L 2018;328:82–209.
[2] European Parliament and Council. Directive (EU) 2019/944 on common rules for the internal market for electricity. Off. J. Eur. Union L 2019;158: 125–99.
[3] Luthander R, Widén J, Nilsson D, Palm J. Photovoltaic self-consumption in buildings: a review. Appl Energy 2015;142:80–94.
[4] Parag Y, Sovacool BK. Electricity market design for the prosumer era. Nat Energy 2016;1:16032.
[5] Morstyn T, Teytelboym A, McCulloch MD. Bilateral contract networks for peer-to-peer energy trading. IEEE Trans Smart Grid 2019;10:2026–35.
[6] Pasqui M, Gerini F, Jacobs M, Carcasci C, Paolone M. Self-dispatching a renewable energy community by means of battery energy storage systems. J Energy Storage 2025;114:115837.
[7] Ramos S, Roque LAC, Gomes A, Soares J, Rolle JC, Vale Z. Multi-objective model for residential energy management in context of individual self-consumption. Math Comput Simul 2025;231:120–7.
[8] Bauwens T, Gotchev B, Holstenkamp L. What drives the development of community energy in Europe? Energy Res Soc Sci 2016;13:136–47.
[9] Creamer E, et al. Community energy: entanglements of community, state, and private sector. Prog Hum Geogr 2018;42:701–21.
[10] Litjens GBMA, Worrell E, van Sark WGJHM. Influence of demand patterns on optimal storage and PV capacities. Sol Energy 2017;155:1276–88.
[11] Barroso LA, et al. Flexibility and P2P energy trading in distribution systems. In: Proceedings of the 2020 IEEE PES general meeting; 2020.
[12] Garey MR, Johnson DS. Computers and intractability: a guide to the theory of NP-completeness. San Francisco: W. H. Freeman; 1979.
[13] Pena-Bello A, Burer M, Patel MK, Parra D. Optimizing PV and grid charging in combined applications. J Energy Storage 2017;13:58–72.
[14] Barbour E, Parra D, Awwad Z, González MC. Community energy storage: a smart choice for the smart grid? Appl Energy 2018;212:489–97.
[15] Tushar W, et al. Peer-to-peer energy systems for connected communities. Appl Energy 2021;282:116131.
[16] Frieden D, Tuerk J, Nabe C, Roberts A. Collective self-consumption and energy communities. Energies 2021;14:1372.
[17] Lazdins R, et al. Synergy between demand flexibility and energy communities: a literature review. Sustainability 2026;18:1858.
[18] Volpato G, et al. General guidelines for the optimal economic aggregation of prosumers in energy communities. Energy 2022;258:124800.
[19] Ferrando M, et al. Urban building energy modeling tools: a review. Sustain Cities Soc 2020;62:102408.
[20] Kwac J, Flora J, Rajagopal R. Household energy consumption segmentation using hourly data. IEEE Trans Smart Grid 2014;5:420–30.
[21] Wang Z, et al. A review of the topology optimization of microgrids. IEEE Access 2019;7:104184–201.
[22] Golari M, Fan N, Wang J. Two-stage stochastic optimal islanding operations. Electr Power Syst Res 2014;114:28–39.
[23] Aliyan E, et al. Optimal formation of energy communities under technical-spatial constraints. In: Proceedings of the 15th international conference on power, energy and electrical engineering (CPEEE). IEEE; 2025.
[24] Sovacool BK, Ratan MH. Sustainable energy transformations, power and politics. Glob Environ Change 2014;29:105–15.
[25] Ministerio para la Transición Ecológica. Real decreto 244/2019 de 5 de abril, Boletín Oficial del Estado no. 83, 35674–35719; amended by Real Decreto-ley 29/2021 de 21 de diciembre, extending the proximity limit from 500 m to 2 km, Boletín Oficial del Estado no. 305, 2019.
[26] Enedis. Guide pratique de l’autoconsommation collective. Tech. rep Paris, France: Enedis; 2022.
[27] Presidenza del Consiglio dei Ministri. Decreto legislativo 8 novembre 2021, n. 199, Gazzetta Ufficiale della Repubblica Italiana no. 285, 2021.
[28] Moreno A, Mello J, Villar J. Comparison of selected self-consumption regulatory approaches in Europe. Heliyon 2025;11(12).
[29] Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of the IEEE international conference on neural networks, vol. 4; 1995. p. 1942–8.
[30] Shi Y, Eberhart RC. A modified particle swarm optimizer. In: Proceedings of the IEEE world congress on computational intelligence; 1998. p. 69–73.
[31] Jain R, Chiu D, Hawe W. A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Tech. Rep DEC-TR-301 Digital Equipment Corporation; 1984.