# Correcting aggregation errors induced by spatial and temporal clustering for MILP optimization of multi-energy systems: a preprocessing approach

*Célia Benmansour[a], Timothé Gronier[a] and Pierrick Haurant[a]*

[a] *IMT Atlantique, GEPEA, UMR CNRS 6144, Nantes, France, pierrick.haurant@imt-atlantique.fr CA*

**Abstract:**
Directly optimizing the sizing and operation of multi-energy systems across building portfolios comprising several hundred buildings over several decades entails prohibitive computational costs. Modelers commonly resort to spatial clustering (aggregating buildings into representative clusters) and temporal clustering (reducing annual time series to representative days). While well-established methods exist for selecting representative days and quantifying temporal errors *a posteriori*, less attention has been devoted to errors arising from spatial aggregation in multi-energy systems mixed-integer linear programming (MILP) optimization. This work identifies three sources of aggregation-induced discrepancies. First, coverage error: spatial aggregation obscures local synchronization between generation and demand, leading to overestimated self-consumption. Second, pooling error: aggregating storage capacities at the cluster level creates inter-node flexibility that does not physically exist, resulting in overestimated storage utilization. Third, peak diversity error: aggregating thermal demand profiles smooths peak loads, causing heating systems to be undersized. The core methodological challenge lies in the endogenous nature of these biases: the solver simultaneously determines capacities and their allocation within each cluster, yet coverage bias depends on the spatial correspondence between generation and demand, unknown *a priori* since it is determined by the optimization itself. We propose a preprocessing module, integrated upstream of the MILP solver, that generates corrective coefficients embedded in the model constraints. Corrections are applied through multivariate affine adjustment for self-consumption rates, storage capacities, and heating power requirements. A simulation grid is generated by varying installed capacities and deployment strategies within clusters. Discrepancies between aggregated and disaggregated models are characterized via multivariate regression, yielding corrective coefficients expressed as functions of decision variables. Applied to a 50-building case study, the correction reduces the cost estimation error from $-23.4\,\%$ to within $-5.6\,\%$ to $-7.5\,\%$, while maintaining a $3\times$ computational speedup.

**Keywords:**
Clustering, MILP optimization, Multi-energy planning, Spatial clustering errors.

# 1. Introduction

The decarbonization of the building sector is a central challenge of the European energy transition. Buildings account for approximately 40% of final energy consumption and 36% of energy-related greenhouse gas emissions in the European Union [1]. Achieving the targets set by the Paris Agreement and the European Green Deal requires a deep transformation of how energy is produced and consumed at the building and district scales [2]. In France, local authorities are key actors of this transformation though institutional instruments.

In this context, multi-energy systems (MES) that couple electrical, thermal, and gas networks at the district scale have emerged as a promising approach. MES exploit synergies between energy carriers and local energy exchanges can reduce grid dependency [3], specifically at urban scale [4,5]. Designing such systems requires long-term investment planning models that determine optimal equipment

capacities and operational strategies over horizons of 20 to 30 years reflecting both equipment lifespans and the time frames set by regulatory decarbonization targets. These models must capture the temporal variability of renewable generation and demand at hourly resolution to accurately represent self-consumption dynamics, storage cycling, and peak demand events. Mixed-Integer Linear Programming (MILP) has become the standard framework for these problems, offering global optimality guarantees and the ability to handle discrete investment decisions [6, 7]. However, the computational cost of MILP grows with the product of spatial nodes, temporal steps, and planning years: both solve time and memory requirements increase rapidly, often exceeding the capacity of standard workstations for portfolios of several dozen buildings over multi-decadal horizons. This creates a tension between the model fidelity needed to capture energy dynamics and the computational tractability required for planning tools. The established approach to address this challenge combines two reduction techniques:

- *Temporal clustering* reduces the 365 days of a year to few representative days, typically selected via K-medoids method applied on demand and weather profiles [8–10]. Recent reviews provide comprehensive analysis of these temporal reduction strategies [11, 12].
- *Spatial clustering* groups, in clusters or energy nodes, buildings of similar characteristics such as demand profiles, equipment, and geographic proximity ; each represented by a single aggregated node in the optimization model [13, 14].

While temporal clustering methods for energy system optimization have been extensively studied [8, 11, 15, 16], spatial clustering has received comparatively less attention. Recent work has quantified spatial aggregation errors in urban energy models, showing that errors can exceed 10–20% at the district scale [17, 18], and some approaches have been proposed to adaptively control spatial resolution [19, 20]. However, most existing works focus on *selecting* good clusters, optimizing the number of clusters, the clustering algorithm, or the feature space rather than on *correcting* the residual errors that persist even with well-chosen clusters. Indeed, while spatial clustering preserves aggregate totals, it destroys the **spatial distribution within clusters**: before clustering, demand and production are known for each building at each timestep, whereas after clustering, only cluster-level quantities are available. This loss of within-cluster information induces errors that can push the optimization in the wrong direction. To the best of our knowledge, no approach has been proposed to correct spatial aggregation errors in multi-energy system MILP models while preserving problem linearity. In that context, the main contributions of this paper are:

- an identification and characterization of three aggregation-induced errors in MILP optimization of multi-energy systems: coverage error (self-consumption overestimation), pooling error (storage capacity overestimation), and peak diversity error (heating system undersizing).
- a preprocessing methodology that computes corrective coefficients through simulation and multivariate regression, and integrates them as affine functions of decision variables directly into MILP constraints, preserving problem linearity.
- a validation on a 50-building case study in France (Nantes), demonstrating that the correction reduces the cost prediction error from $-31.7\%$ to $-18.1\%$ (i.e. the aggregated model predicts the true deployed cost within 18% instead of 32%), reduces the MILP cost error from $-23.4\%$ to within $-5.6\%$, while maintaining a $3\times$ computational speedup.

# 2. Methodology

In its simplest form, the optimization flow consists of two stages: raw input data (demand profiles, weather data, technology parameters for each of the $N$ buildings, over $Y$ years of $T$ hourly annual timesteps) feed directly a MILP solver, which determines optimal investments and operational dispatch at each timesteps. However, the resulting problem size quickly becomes computationally intractable for municipal-scale building portfolios, typically comprising several hundred to several

thousand buildings and multi-decadal planning horizons, both in terms of solve time and memory requirements.

To reduce the problem dimension, modelers introduce two clustering stages upstream of the solver. Spatial clustering aggregates the $N$ buildings into $K \ll N$ representative clusters, each described by a single aggregated demand and production profile. Temporal clustering reduces the 365 annual days to $R_D \ll 365$ representative days: reduces the $T$ annual timesteps to $R_D \times H$ hourly steps. Together, these reductions shrink the optimization problem to enabling tractable solve times at the cost of the aggregation errors identified in Section 1.. The contribution of this work is the addition of a **preprocessing module** positioned *between* the clustering step and the MILP optimization (Figure 1). This module takes as input the clustered data together with the original disaggregated profiles, computes corrective coefficients ($\beta$ and $\alpha$) through simulation, and injects them as constants into the MILP constraints.

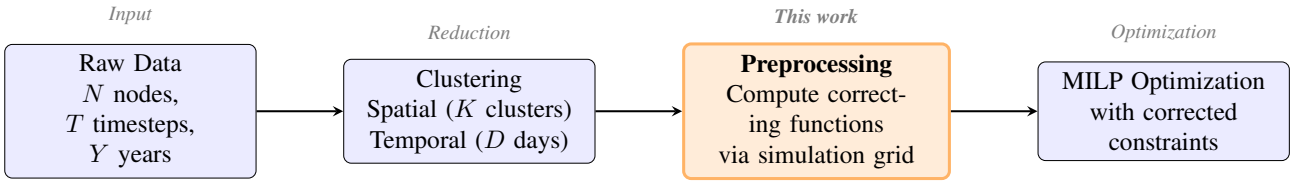


**Figure 1**: Position of the preprocessing module in the optimization pipeline.

## 2.1. Characterization of errors induced by spatial clustering

In the French regulatory framework, two distinct self-consumption schemes are distinguished. **Individual self-consumption** (ISC) refers to energy produced and consumed within the same building or legal entity, behind a single metering point. **Collective self-consumption** (CSC), enabled at the European level by the revised Renewable Energy Directive [21, 22], allows energy produced by one building to be shared with others within a defined perimeter, subject to network constraints and regulatory limits. This distinction is critical for optimization models because ISC and CSC are governed by different tariff structures, taxation rules, and grid access conditions, meaning that misattributing energy flows between the two schemes leads to incorrect economic evaluation of investments. CSC fundamentally alters optimal sizing and dispatch strategies: surplus generation at one node can offset demand at another, creating inter-node dependencies that the optimization model must capture explicitly [23].

When buildings within a cluster have different generation and demand patterns, the aggregated model implicitly assumes that surplus energy from one building can offset the deficit of another, even without any physical connection. This "fictitious flexibility", called **coverage error** in this paper, leads to overestimation of self-consumption and underestimation of grid dependency, since the sum of individual self-consumed quantities is always less than or equal to the self-consumption computed from aggregated profiles [23, 24]. When spatial clustering is applied to systems involving both ISC and CSC, merging multiple buildings into a single equivalent node, blurs the boundary between ISC and CSC: energy that would physically remain behind an individual meter (ISC) becomes indistinguishable from energy shared across buildings (CSC). The model thus conflates physical energy sharing through CSC with fictitious sharing introduced by aggregation, making it impossible to correctly allocate self-consumed energy to the appropriate regulatory scheme and leading to distorted investment signals.

Now, consider three buildings A, B and C within a cluster at a given timestep $t$. Each building's $ISC_n$ is bounded by the minimum of its local production $P_n(t)$ and its local demand $D_n(t)$ :

$$ISC_n(t) = \min\left(P_n(t),\ D_n(t)\right) \tag{1}$$

**Table 1**: Disaggregated view: self-consumption is computed building by building.

| Node $n$ | $D_n(t)$ [kW] | $P_n(t)$ [kW] | $ISC_n(t)$ [kW] | Surpl. [kW] |
|---|---|---|---|---|
| A | 12 | 0 | 0 | 0 |
| B | 3 | 8 | 3 | 5 |
| C | 2 | 5 | 2 | 3 |
| $\sum_n$ | **17** | **13** | **5** | **8** |
| Cluster | **17** | **13** | **13** | – |

After aggregation, the optimizer sees a single cluster with 17 kW demand and 13 kW production, yielding 13 kW of self-consumption: a **160 % overestimation** compared to the real 5 kW $ISC_n$; and the 8 kW of real surplus vanishes entirely (Table 1). This occurs because the aggregated model implicitly assumes that building B's surplus can offset building A's deficit, even though no physical connection exists between them. Formally and generally, for a cluster $c$:

$$ISC_c = \sum_{n \in c} ISC_n = \sum_{n \in c} \min(P_n(t),\, D_n(t)) \leq \min\left(\sum_{n \in c} P_n(t),\, \sum_{n \in c} D_n(t)\right) \tag{2}$$

To quantify the discrepancy presented by Eq. (2), we introduce a **coverage coefficient** $\beta_c^{cov}(t) \in [0, 1]$, that measures the fraction of the aggregated self-consumption that is physically achievable for each cluster $c$ and timestep $t$ by :

$$\beta_c^{cov}(t) = \frac{\sum_{n \in c} \min\left(P_n(t),\, D_n(t)\right)}{\min\left(\sum_{n \in c} P_n(t),\, \sum_{n \in c} D_n(t)\right)} \tag{3}$$

$\beta_c^{cov} = 1$ means no difference; $\beta_c^{cov} < 1$ means the aggregated model overestimates self-consumption. In the example of Table 1, $\beta_c^{cov} \approx 0.38$.

Besides, the aggregated model creates a "virtual pool" of storage capacity, implying that surplus energy at one building can be stored in another building's battery or thermal storage. In reality, a storage device can only store local surplus energy. This overestimation of effective storage, named **pooling error** in this paper, reduces the incentive to invest in appropriately sized storage systems [25]. Continuing the previous example, consider the surplus distribution and storage installations (Table 2) :

Only 2 kWh can actually be stored, indeed, building A has storage but no surplus, building B has surplus but no storage, and only building C has both. After aggregation, the model sees an 8 kW surplus and 7 kWh storage capacity, perceiving 7 kWh storable, a **250 % overestimation** (Table 2). Moreover, if the aggregated model perceives less surplus than actually exists, the storage error is further distorted. Because storage introduces *temporal coupling* through its state of charge, this error does not remain confined to a single timestep: an overestimated charge at time $t$ inflates the available energy at $t + \Delta t$, which in turn affects discharge decisions, grid exchanges, and subsequent charging opportunities. The pooling error thus propagates across the entire optimization horizon, accumulating through the state-of-charge dynamics and potentially distorting operational strategies far beyond the timesteps where the original mismatch occurs. The pooling coefficient $\beta_c^{pool}$ follows the same structure as $\beta_c^{cov}$ but is applied to storage charge and discharge flows.

**Table 2**: Disaggregated view: storage requires co-location of surplus and storage capacity.

| **Node** | **Surp.** [kW] | **Storage Capacity** [kWh] | **Stored** [kWh] |
|---|---|---|---|
| A | 0 | 5 | 0 |
| B | 5 | 0 | 0 |
| C | 3 | 2 | 2 |
| $\sum_n$ | **8** | **7** | **2** |
| Cluster | **8** | **7** | **7** |

**Table 3**: Non-coincident heating demand ($D_n^{\text{heat}}$) peaks across three buildings.

| | $D_n^{heat}(t)$ [kW] | | | |
|---|---|---|---|---|
| **Node/$t$** | **7 am** | **8 am** | **9 am** | **Peak** |
| A | 15 | 13 | 8 | 15 |
| B | 3 | 11 | 12 | 12 |
| C | 2 | 8 | 10 | 10 |
| $\sum_n D_n(t)$ | 20 | **32** | 30 | **37** |

Finally, thermal demand peaks in different buildings do not necessarily occur at the same time. When individual profiles are aggregated, these non-simultaneous peaks partially cancel out, producing a smoothed cluster peak that is lower than the sum of individual peaks, defined as **Peak diversity error** in this paper. Since heating systems must be sized for each building's own peak demand, the aggregated model underestimates total installed heating capacity. This effect, well known in electrical engineering as the *diversity factor*, has been documented for heating loads with reported peak load reductions of 15–47% when aggregating residential buildings [26, 27]. Consider three buildings with non-simultaneous thermal peaks (Table 3), each building requires a heating system sized for its individual peak.
Individual peaks occur at various hours so that the aggregated profile never reaches their sum : the sum of individual peaks is 37 kW, while the aggregated demand profile reaches a simultaneous maximum of only 32 kW at 8:00, when none building is at its own peak. An aggregated model therefore sizes heating at 32 kW instead of 37 kW, a **14 % underestimation**. The ratio between individual building peaks which may occur at different times and the simultaneous cluster peak, defines as a **peak diversity factor**, captures this effect:

$$\alpha_c = \frac{\sum_{n \in c} \max_t D_n^{heat}(t)}{\max_t \sum_{n \in c} D_n^{heat}(t)} \geq 1 \tag{4}$$

$\alpha_c = 1$ means all buildings peak simultaneously ; $\alpha_c > 1$ means peaks are diversified and the aggregated peak underestimates required capacity. Unlike $\beta_c^{cov}$ and $\beta_c^{pool}$, $\alpha_c$ depends only on the demand profiles and does not depend on investment decisions. It is therefore computed once per cluster from the original demand data.

## 2.2. Preprocessing module: corrective coefficients for MILP constraints

Precalculated coefficients $\beta$ and $\alpha$, obtained from raw data, are embedded into the MILP constraints as **constants** (from preprocessing), while only capacity and power variables remain as decision variables, so that all corrections preserve constraint linearity.
The self-Consumption constraint limits the self-consumed power to the fraction that is physically achievable given the spatial distribution of production and demand within the cluster. Let $P_{c,t}^{sc}$ denote the self-consumed production power, $C_c^{prod}$ the installed production capacity and $D_{c,t}$ demand at cluster $c$ and timestep $t$. The original self-consumption constraint is corrected by $\beta_c^{cov}(t)$:

$$P_{c,t}^{sc} \leq \beta_c^{cov}(t) \cdot D_{c,t} \tag{5}$$

$\beta^{cov}$ dependency on installed production capacity is modelled as a linear regression obtained by ordinary least squares method on the samples generated by an experimental design :

$$\beta_c^{cov}(t) = \beta_{0,c}^{cov}(t) + \gamma_c^{cov}(t) \times \frac{C_c^{prod}}{C_{ref,c}^{prod}} \tag{6}$$

Besides, let $P_{c,t}^{ch}$ and $P_{c,t}^{\text{dis}}$ denote storage charge and discharge powers, $C_c^{stor}$ the installed storage capacity, and $S_{c,t}^{sur}$ the surplus production available for storage. The corrected storage constraints are given by :

$$P_{c,t}^{ch} \leq \beta_c^{pool,ch}(t) \times S_{c,t}^{sur} \tag{7}$$

$$P_{c,t}^{dis} \leq \beta_c^{pool,dis}(t) \times D_{c,t} \tag{8}$$

Storage systems add complexity because their utilization depends on **two capacities simultaneously**: production capacity, which determines where surplus exists ; and storage capacity which determines where surplus can be absorbed, the pooling coefficient for storage charging is modelled as a bivariable regression :

$$\beta_c^{pool,ch}(t) = \beta_{0,c}^{ch}(t) + \gamma_{prod,c}^{ch}(t) \cdot \frac{C_c^{prod}}{C_{ref,c}^{prod}} + \gamma_{stor,c}^{ch}(t) \cdot \frac{C_c^{stor}}{C_{ref,c}^{stor}} \tag{9}$$

The regression procedure extends to a 2D capacity grid over both production and storage dimensions, with the same deployment strategies applied to each. An analogous formulation applies for discharge coefficients $\beta_c^{pool,dis}(t)$. These constraints apply identically to any production-storage pair (e.g., PV and electrical battery, or heat pump and thermal energy storage), with the appropriate capacity and power variables substituted for each technology.
Finally, let $C_c^{\text{heat}}$ denote the total installed heating capacity at cluster $c$ (summing all heating technologies). The heating capacity constraint becomes:

$$C_c^{heat} \geq \alpha_c \cdot \max_t D_{c,t}^{heat} \tag{10}$$

This ensures that each cluster's heating system is sized for the sum of individual building peaks, not the smoothed aggregated peak. $\alpha_c$ depends only on demand profiles and is computed once per cluster from the original data using Eq. (4). Since $\alpha_c$ is a constant, linearity is preserved.

## 2.3. The distribution capacity dependency problem

The errors introduced by spatial aggregation are *endogenous*: they depend on optimization results (installed capacities, dispatch patterns) that are themselves affected by the aggregation. This circular dependency makes *post-hoc* correction insufficient and calls for corrections that operate within the optimization loop. The coverage and pooling coefficients, resp. $\beta^{cov}$ and $\beta^{pool}$, depend on **how capacities are distributed across buildings within clusters**. Consider the coverage example in Table 1: if the total production capacity were allocated differently among buildings A, B, and C, the mismatch between local production and demand would change, yielding a different value of $\beta^{cov}$. Formally:

$$\beta(t) = f(\mathbf{C}^*) \qquad \beta \text{ coefficients depend on optimal capacities } \mathbf{C}^* \tag{11}$$

$$\mathbf{C}^* = \arg\min_{\mathbf{C}} F(\beta^\bullet) \qquad \text{optimal capacities } \mathbf{C}^* \text{ depend on coefficients } \beta \tag{12}$$

This circular dependency is compounded by the multi-year planning structure of the optimization. The MILP model determines investments $I_{s,c,y}$ for each technology $s$, cluster $c$, and year $y$ over the full planning horizon. At each year, the optimizer updates capacity to a cluster as a whole, without specifying how this capacity should be distributed among the individual buildings within that cluster. Yet the corrective coefficients precisely quantify the error introduced by this unknown intra-cluster distribution. Since the preprocessing module runs *before* the optimization, it must anticipate the bias for capacity levels that the solver has not yet determined and that will evolve across multiple investment periods.
A first glance approach to address this issue would require iterative optimization, recomputing coefficients after each solve, then re-solving with updated values negating the computational gains from clustering. The following section presents the simulation-based approach adopted to break this circularity by precomputing coefficients over a grid of plausible capacity levels, so that the MILP constraints can interpolate the appropriate correction at any optimization variables chosen by the solver.
Rather than solving the fixed-point problem, we explore the space of possible capacity distributions through predefined **deployment strategies**. Each strategy $s$ defines a rule for distributing a given total cluster capacity $C_c$ among individual buildings $n \in c$ where the system studied could be installed. Five strategies are considered, spanning from economically rational to deliberately adverse configurations. *Uniform* assigns equal capacity to every building regardless of size (Eq. (13)); *Proportional* weights capacity by annual energy demand, so that larger consumers receive larger installations (Eq. (14)); *Concentrated* allocates a fraction $\phi$=0.8 of the cluster capacity to the top 20% consumers (by demand), and spreads the remainder among the rest (Eq. (15)) ; *Dispersed* inverts the proportional rule, weighting by inverse demand; this represents a worst-case scenario where installations are mismatched with consumption (Eq. (16)) ; *Random* draws uniform weights $\xi_n \sim \mathcal{U}(0,1)$ to capture deployment uncertainty (Eq. (17)). This bracketing approach allows assessing the robustness of

the correction: if the coefficients reduce errors even under worst-case deployment assumptions, their effectiveness under realistic conditions is ensured.

$$C_n^{unif} = \frac{C_c}{\mathrm{Card}(c)} \tag{13}$$

$$C_n^{prop} = C_c \cdot \frac{D_n^{tot}}{\sum_{m \in c} D_m^{tot}} \tag{14}$$

$$C_n^{conc} = \begin{cases} C_c \cdot \phi / n_{top} & \text{if } n \in top_{20\%} \\ C_c \cdot (1-\phi)/n_{rest} & \text{otherwise} \end{cases} \tag{15}$$

$$C_n^{disp} = C_c \cdot \frac{1/D_n^{tot}}{\sum_{m \in c} 1/D_m^{tot}} \tag{16}$$

$$C_n^{rand} = C_c \cdot \frac{\xi_n}{\sum_{m \in c} \xi_m}, \quad \xi_n \sim \mathcal{U}(0,1) \tag{17}$$

For each strategy, the total cluster Capacity is varied across a range of levels relative to a reference capacity $C_{ref,c} = \bar{D}_c$ (the mean demand of the cluster). This yields a structured simulation grid that samples the space of possible within-cluster capacity distributions.

# 3. Case study

The proposed methodology is evaluated on a portfolio of 50 tertiary buildings located in France. The portfolio is representative of a medium-sized municipal building stock and includes offices, schools, swimming pools, gymnasiums, museums, libraries, and town halls.

## 3.1. MILP model specification

The case study is formulated as a MILP investment and operation model for a multi-energy building portfolio. The formulation follows classical district energy system models [6, 7], extended with the corrective constraints introduced in Section 2.2.. The techno-economic modeling itself is standard; the methodological contribution lies in the integration of aggregation-correction coefficients within clustered spatial nodes.

The planning horizon is set to $Y = 20$ years. Spatial resolution depends on the scenario: $N = 50$ in the fully disaggregated reference case (S0-T) and $K = 5$ in clustered scenarios (S1 and S2). Temporal reduction is performed using $R_D = 12$ representative days (Section 3.3.).

### 3.1.1. Objective function

The objective minimizes total discounted system cost (Eq. (18)):

$$\min_{\mathbf{x}} F(\mathbf{x}) = \sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}} \sum_{y \in \mathcal{Y}} \frac{CAPEX_s \cdot I_{s,c,y}}{(1+r)^y} + \sum_{c \in \mathcal{C}} \sum_{t \in \mathcal{T}} w_t \cdot C_{c,t}^{op} \tag{18}$$

where $\mathcal{S} = \{PV, Bat, HP, TES\}$ and $r = 4\%$. Operational costs include electricity imports and gas consumption. Given the 20-year horizon, investment costs are accounted as discounted lump-sum expenditures in the year of installation.

### 3.1.2. Energy balance constraints

The main constraints of the MILP model concern electrical and thermal balances (Resp. Eq. (19) and Eq. (20)) and the management of PV surplus, allocated to battery charging, CSC export, or grid curtailment (Eq. (21)).

**Electrical balance.** For each cluster $c$ and timestep $t$:

$$P^{PV,sc}_{c,t} + P^{dis}_{c,t} + P^{grid}_{c,t} + P^{CSC}_{c,t} = D^{elec}_{c,t} + P^{ch}_{c,t} + P^{HP}_{c,t} \tag{19}$$

$$Q^{HP}_{c,t} + Q^{GB}_{c,t} + Q^{TES,dis}_{c,t} = D^{heat}_{c,t} + Q^{TES,ch}_{c,t} \tag{20}$$

$$P^{PV,sur}_{c,t} = P^{ch,PV}_{c,t} + P^{CSC,sent}_{c,t} + P^{curt}_{c,t} \tag{21}$$

**PV surplus balance.** PV surplus (generation exceeding local demand) is allocated to battery charging, CSC export, or grid curtailment:

**Thermal balance.**

### 3.1.3. Technology models

The technology models (PV, battery, TES, heat pump, gas boiler) are formulated as standard linear components [6, 7].

## 3.2. Building portfolio description

Seven building categories are represented (Table 4). Annual electrical and thermal demands follow log-normal distributions calibrated to reflect the dispersion observed in municipal portfolios.

**Table 4**: Physical characteristics, energy systems and investment eligibility of the building portfolio by category

| **Category** | $N$ | $\tilde{E}_{\text{el}}$ [MWh/yr] | $\tilde{E}_{\text{th}}$ [MWh/yr] | **Existing heating** Gas | HP | $\sigma_E$[a] | $f_{\text{PV}}$ [%] | $p_{\text{PV}}$ [%] | $\tilde{A}_{\text{roof}}$ [m$^2$] | Cool. |
|---|---|---|---|---|---|---|---|---|---|---|
| Offices | 25 | 100 | 120 | 15 | 10 | $25-65$ | 88 | 0.8 | 200 | ✓ |
| Schools | 10 | 80 | 150 | 8 | 2 | $40-80$ | 80 | $0.7-0.8$ | 600 | ✓[b] |
| Swimming pools | 4 | 300 | 800 | 4 | 0 | $30-70$ | 75 | 0.5 | 800 | — |
| Gymnasiums | 3 | 120 | 200 | 3 | 0 | $50-85$ | 67 | $0.6-0.7$ | 700 | — |
| Museums | 3 | 200 | 250 | 2 | 1 | $15-40$ | 33 | 0.6 | 250 | ✓ |
| Libraries | 2 | 150 | 180 | 1 | 1 | $45-75$ | 50 | $0.5-0.6$ | 500 | ✓ |
| Town halls | 3 | 180 | 220 | 2 | 1 | $10-30$ | 0 | $0.5-0.6$ | 300 | ✓ |
| **Total** | **50** | | | **35** | **15** | | | | **$\approx$ 37**[c] | **30** |

$f_{\text{PV}}$: PV-usable roof fraction ; $p_{\text{PV}}$: PV eligibility probability
[a]$\sigma_E$ denote shape parameters ; [b]20 % of schools only ; [c]Expected value; actual count is stochastic.

Hourly electrical demand profiles are generated from category-specific occupancy schedules, with three levels: peak morning, peak afternoon, and standard occupied hours. Outside occupation hours, demand drops to a base load ratio drawn uniformly per building. Weekend and school vacation reduction factors are applied per category. A random schedule shift ($\pm$1–2 hours) is applied to each building to introduce non-simultaneity. Annual demand is calibrated by normalizing hourly profiles so that their integral matches the target annual consumption.
Hourly thermal demand profiles are generated using a first-order AutoRegressive and eXogen (ARX) model (Eq. (22)) developed by Akajoule[1]:

$$Q(t) = \max\Big(0,\ \ \varphi\, Q(t-1) + U_{th}\ \max\big(0,\ T_b - T_{ext}(t)\big) + \varepsilon(t)\Big) \tag{22}$$

where $\varphi \in [0, 1]$ is the thermal inertia coefficient, $U_{th}$ [kW K$^{-1}$] the heat loss coefficient, $T_b$ and $T_{ext}$ [K] are base and ambiant temperatures, and $\varepsilon(t) \sim \mathcal{N}(0,\ (0.5\,|U_{th}|)^2)$ a white noise term scaling with $U_{th}$. Variable setpoints (occupied, setback, frost protection) are handled via an effective temperature

[1]Akajoule SAS, Nantes, France. `https://akajoule.com`

transformation. The inertia $\varphi$ and loss coefficient $U_{th}$ are calibrated per building to match target annual demand, producing realistic profiles with temporal autocorrelation, weather dependency, and inter-building diversity.
Five buildings participate in a regulated collective self-consumption scheme. One large office acts as producer, while four offices act as consumers without PV eligibility.

## 3.3. Clustering configuration

To reduce the computational complexity of the MILP, we apply two independent procedures, spatial and temporal clustering, whose parameters are summarized in Table 5. The values $K$=5 and $R_D$=12 are selected as accuracy computation trade-offs, as demonstrated by the sensitivity analysis in Section 4.3..

**Table 5**: Clustering configuration.

| | **Spatial** | **Temporal** |
|---|---|---|
| Algorithm | K-Means [28] | K-medoids [29] |
| Initialization | K-Means++ [30] | K-medoids++ |
| Distance | Euclidean, z-score norm. | Euclidean, z-score norm. |
| Resolution | $K = 5$ clusters | $R_D = 12$ representative days |
| Features | $n_f = n_d + n_e = 7{+}9 = 16$ | $n_{\text{var}} \times H = 4 \times 24 = 96$ |

**Spatial clustering.** Each building $b$ is described by a feature vector comprising two categories: (i) *demand statistics* ($n_d$=7): peak, mean, standard deviation, and load factor for electrical demand; peak and mean for thermal demand; and heat-to-electricity ratio; (ii) *equipment capacities* ($n_e$=9): existing and maximum PV area, heat pump, gas boiler, battery, and TES capacities, grid import/export limits, and cooled surface area. All features are z-score normalized so that each contributes equally to the Euclidean distance ($n_f$=16).
The clustering proceeds in two stages. First, buildings linked by collective self-consumption (CSC) contracts must be grouped within the same optimization node, since they physically exchange energy. In the present case study, one CSC group of 5 buildings forms a single imposed cluster ($K_{\text{CSC}}$=1). Second, the remaining $N'$=$N - 5 = 45$ buildings are partitioned into $K'$=$K - K_{\text{CSC}} = 4$ free clusters by K-Means with K-Means++ initialization [28, 30]. The resulting compact clusters maximize intra-cluster homogeneity of energy profiles, which directly reduces the aggregation bias that the $\beta/\alpha$ corrections must compensate.
After clustering, each cluster is represented by a single aggregated node in the MILP. Extensive quantities (demands, installed capacities) are summed across cluster members. Intensive quantities are aggregated via capacity-weighted means: equipment efficiencies and unit investment costs are weighted by installed capacity (e.g., gas boiler efficiency by nominal boiler power), while energy prices are weighted by electrical demand.

**Temporal clustering.** Each day $j$ is characterized by concatenating the 24-hour profiles of four variables: territory-level electrical demand (summed across all buildings), territory-level thermal demand, outdoor temperature, and global horizontal irradiance. K-medoids with K-medoids++ initialization [29, 31] selects actual observed days as medoids rather than virtual centroids, preserving inter-variable correlations (e.g., between irradiance and demand) that would otherwise be smoothed out, distorting peak demand events and investment sizing [8, 16]. Each representative day $d$ receives a weight $w_d$ equal to its cluster cardinality ($\sum_d w_d = 365$), used to scale the MILP objective. For multi-year horizons, the same $R_D$ representative days and their weights are replicated identically for each planning year, assuming stationary climate conditions.

## 3.4. Validation protocol

The preprocessing module computes coefficients through a structured experimental design. For each cluster $c$, a reference capacity $C_{ref,c}$ is defined as the **maximum installable capacity**, obtained by summing the physical upper bounds of all buildings within the cluster (e.g., available roof area for PV, maximum battery parameter for storage). This anchors the regression grid to the MILP decision domain $[0, C_c^{\max}]$. Capacity is varied across 7 levels uniformly spaced in $[0, C_{ref,c}]$. For 1D regressions ($\beta_c^{cov}$), this yields 7 samples per cluster and timestep. For 2D regressions ($\beta_c^{pool,ch}$), a grid of $7 \times 7 = 49$ capacity combinations is evaluated.

Coefficients are computed for each **clustered timestep** $t \in \{0, \ldots, R_D \times H - 1\}$. With $R_D = 12$ representative days and $H = 24$ hours, this yields 288 coefficient values per cluster, a key distinction, since the same calendar hour on a winter day exhibits different bias characteristics from a summer day.Three scenarios are compared: **S0-T** (nodal reference, $K$=$N$=50, temporal clustering only, $R_D$=30); **S1** (spatial and temporal clustering without correction, $K$=5, $R_D$=12); and **S2** (same clustering with $\beta/\alpha$ correction, $K$=5, $R_D$=12). The reference scenario uses a higher temporal resolution ($R_D$=30) to maximize fidelity at full spatial resolution. Each S2 run uses a **single deployment strategy** to distribute capacity within clusters during the preprocessing simulation, producing strategy-specific corrective coefficients that are then embedded in the MILP. Since $\beta$ depends on how capacity is distributed, different strategies yield different coefficients and therefore different optimal solutions. Five strategies are tested independently.

Two error metrics are defined. The **aggregation error** $\varepsilon_{agg}$ isolates the dispatch bias introduced by spatial clustering: S0-T optimal capacities are fixed in the aggregated model, and the resulting cost is compared to the S0-T reference cost. This metric captures only the operational approximation error, independent of capacity sizing. The **prediction gap** $\varepsilon_{pred}$ measures the end-to-end reliability of the aggregated model as a planning tool. It compares the optimization cost $Cost_{opt}$ (the Gurobi objective on the aggregated model) against the deployed cost $Cost_{dep}$ (obtained by disaggregating the cluster-level capacities back to individual buildings using the chosen deployment strategy, and performing a full 50-building dispatch):

$$\varepsilon_{agg} = \frac{Cost_{agg} - Cost_{S0\text{-}T}}{Cost_{S0\text{-}T}} \tag{23}$$

$$\varepsilon_{pred} = \frac{Cost_{opt} - Cost_{dep}}{Cost_{dep}} \tag{24}$$

A negative $\varepsilon_{pred}$ indicates that the aggregated model underestimates the true system cost, meaning the decision-maker expects lower costs than will actually be realized once capacities are deployed. This metric is the most relevant for investment decisions, as it quantifies the gap between what the optimizer promises and what the system delivers.

# 4. Results and discussion

The framework is applied to the case study described in Section 3.1. over a 20-year planning horizon. Since the same representative days are replicated identically for each planning year under stationary climate assumptions, the preprocessing module computes corrective coefficients for a single year and replicates them across the horizon. In the reference scenario (S0-T), total electricity demand amounts to 8.2 GWh/yr and thermal demand to 14.3 GWh/yr across the 50-building portfolio. Electricity is supplied by grid imports (67 %), PV self-consumption (24 %), and battery discharge (9 %), with a self-consumption rate of 71 %. Thermal demand is covered almost equally by heat pumps (48 %) and gas boilers (48 %), with marginal TES contribution (4 %).

## 4.1. Scenario comparison, error decomposition, and cost analysis

**Table 6**: Capacity sizing and cost prediction across scenarios and deployment strategies.

| Scenario | Strategy | PV [$kW_p$] | Battery [kWh] | HP [$kW_{th}$] | TES [$kWh_{th}$] | Total Cost [k€] | Solve [s] | Depl. Cost [k€] | $\varepsilon_{pred}$ [%] |
|---|---|---|---|---|---|---|---|---|---|
| S0-T (Ref.) | — | 11,095 | 7,952 | 2,729 | 5,416 | 49,060 | 181 | — | — |
| S1 (No corr.) | — | 12,786 (15.2 %) | 7,985 (0.4 %) | 4,101 (50.3 %) | 9,085 (67.7 %) | 37,586 (–23.4 %) | 21 ×8.6 | 55,019 | −31.7 |
| S2 (Corrected) | Uniform | 12,208 (10.0 %) | 7,924 (–0.4 %) | 2,806 (2.8 %) | 396 (–92.7 %) | 45,758 (–6.7 %) | 56 | 55,872 | −18.1 |
| | Proportional | 12,153 (9.5 %) | 7,924 (–0.4 %) | 2,911 (6.7 %) | 396 (–92.7 %) | 45,400 (–7.5 %) | 54 | 55,735 | −18.5 |
| | Concentrated | 12,251 (10.4 %) | 7,875 (–1.0 %) | 2,634 (–3.5 %) | 7,655 (41.3 %) | 46,318 (–5.6 %) | 60 | 59,468 | −22.1 |
| | Dispersed | 11,772 (6.1 %) | 7,875 (–1.0 %) | 2,749 (0.7 %) | 6,903 (27.5 %) | 46,333 (–5.6 %) | 60 | 57,099 | −18.9 |
| | Random | 12,086 (8.9 %) | 7,911 (–0.5 %) | 2,676 (–1.9 %) | 7,855 (45.0 %) | 46,311 (–5.6 %) | 60 | 57,293 | −19.2 |

Table 6 reports the results. Without correction (S1), aggregation overestimates PV by 15.2 % and HP by 50.3 %, while battery remains within 1 % of the reference. TES is severely overestimated (67.7 %), reflecting the mismatch between aggregated and nodal storage needs. The optimization cost is underestimated by –23.4 %, yielding a prediction gap of –31.7 %: the model predicts a cost 32 % below reality. The $\beta/\alpha$ correction (S2) reduces the cost error to within –5.6 to –7.5 % and significantly narrows the prediction gap to –18.1 % (uniform). HP sizing improves from 50.3 % to –3.5–6.7 % thanks to the combined action of $\alpha$ and $\beta_{HP}$; PV errors decrease to 6.1–10.4 %; battery corrections remain at most 1.0 %. TES sizing is strongly strategy-dependent: uniform and proportional yield near-zero investment (–92.7 %) because their $\gamma$ coefficients collapse to zero, whereas concentrated (41.3 %) and random (45.0 %) produce steeper slopes that restore significant TES investment, albeit with substantial oversizing relative to the reference (Section 4.2.). The aggregation error $\varepsilon_{agg}$=–0.8 %, confirming that the aggregated dispatch itself introduces only a modest cost underestimation. The prediction gaps are far larger : $|\varepsilon_{pred}|$=18–32 %, demonstrating that the dominant error source is capacity sizing and spatial redistribution, not the dispatch approximation. Deployed cost decomposition reveals that S2 achieves lower CAPEX than S0-T through mutualization, but total deployed costs exceed S0-T due to OPEX penalties from imperfect capacity redistribution within clusters. When the optimizer assigns a cluster-level HP capacity that is then redistributed across individual buildings, some buildings inevitably receive less than their optimal share. For those under-endowed buildings, gas backup compensates the shortfall at a disproportionately high marginal cost, while over-endowed buildings merely waste idle CAPEX. Since the gas penalty grows faster than the CAPEX saving, the net effect is a deployed cost above S0-T. This asymmetry explains why S1, despite large capacity errors, achieves the lowest deployed cost (55.019 k€): its systematic HP oversizing leaves fewer buildings under-endowed, inadvertently limiting gas backup exposure. It also explains why the prediction gap, not the deployed cost, is the relevant metric for investment decisions. Among S2 strategies, uniform minimizes $|\varepsilon_{pred}|$ (–18.1 %) while concentrated yields the largest gap (–22.1 %) due to severe spatial mismatches.

## 4.2. Corrective coefficient analysis

**PV self-consumption ($\beta^{cov}$).** Figure 2 and 3 present spatio-temporal structures of $\beta_0^{cov}(d,h)$ and $\gamma^{cov}(d,h)$ for Cluster 0. Corrections are active only during solar hours (6 am to 9 pm), with summer days exhibiting the strongest corrections ($\beta_0$ down to $0.50$, $\gamma$ reaching $-0.70$) due to a higher PV surplus amplifying the production–demand mismatch. Inter-cluster variation is substantial (Figure 3): Cluster 2 (20 buildings) requires the strongest correction ($\bar{\beta}_{sc}\approx 0.83$ at full capacity, meaning 17 % of

aggregated self-consumption is fictitious), while Cluster 4 requires the mildest ($\gamma = -0.033$). The correction magnitude correlates with intra-cluster heterogeneity in demand profiles and PV eligibility ratios.

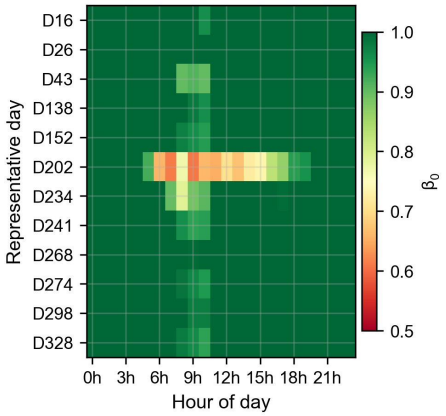


**Figure 2**: PV self-consumption correction : Spatio-temporal coefficients $\beta_0(d, h)$ for Cluster 0

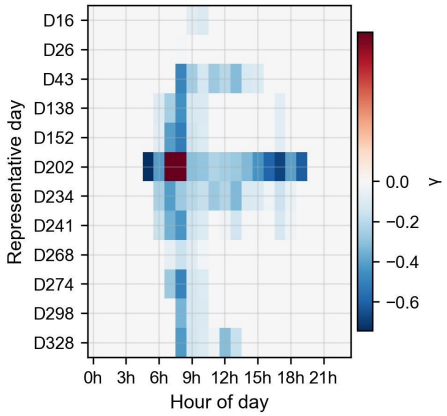


**Figure 3**: PV self-consumption correction : Spatio-temporal coefficients $\gamma(d, h)$ for Cluster 0.

**Battery storage ($\beta_{Bat}^{pool}$).** The battery charge correction surface (Cluster 2) reveals a narrow correction range (0.952–1.000): the negative PV slope ($\gamma_{PV} = -0.035$) reflects surplus generation at buildings without batteries, while the positive battery slope ($\gamma_{Bat} = +0.014$) captures increased co-location probability. Although modest per timestep, these corrections accumulate over the 20-year horizon and explain the $\leq 1\%$ battery sizing errors in Table 6.

**HP thermal saturation ($\beta_{HP}$).** Figure 5 reveals two distinct correction regimes. Clusters 0 and 2 ($\beta_0 \approx 0.93$, $|\gamma| < 0.10$) exhibit mild corrections with relatively synchronized thermal peaks. Clusters 1 and 4 ($\beta_0 \approx 0.65$, $|\gamma| \approx 0.67$) exhibit severe corrections: $\beta_{HP}$ drops below 0.05 at full capacity, meaning nearly all aggregated HP output is fictitious. These clusters group buildings with highly diverse thermal behaviors (swimming pools alongside offices). At the global level, $\beta_0 = 0.659$ and $\gamma = -0.514$ : at 70 % of reference capacity, only 30 % of the aggregated HP output is physically realizable. This correction, combined with $\alpha$, reduces HP overestimation from 50 % to $-3.5$–$6.7$ %.

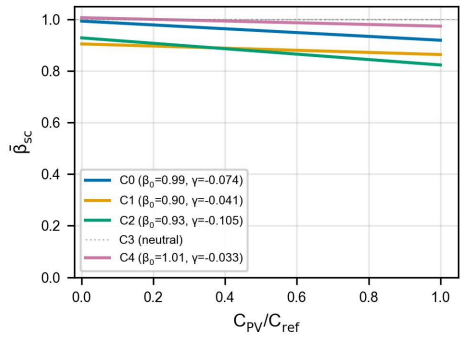


**Figure 4**: Time-averaged $\bar{\beta}_{sc}$ vs. normalized PV capacity per cluster

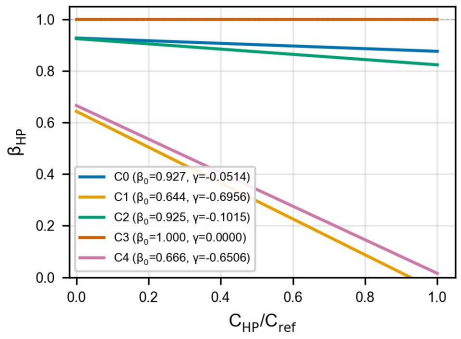


**Figure 5**: ime-averaged HP saturation coefficient $\beta_{HP}$ vs. normalized HP capacity per cluster.

**TES storage ($\beta_{TES}^{pool}$).** The TES charge correction (Cluster 1) shows a 3% overestimation at the intercept ($\beta_0 = 0.970$), with $\gamma_{HP} = -0.070$ and $\gamma_{TES} = +0.070$, while the discharge correction is milder ($\beta_0 = 0.982$, range 0.960–0.981). However, these values mask a structural identifiability issue: for uniform and proportional strategies, the $\gamma$ coefficients collapse to near zero ($|\gamma| < 0.001$), yielding flat correction surfaces that provide no penalty signal, hence the –92.7 % TES underinvestment in

Table 6. Concentrated and random strategies generate the asymmetric distributions needed for meaningful slopes, restoring significant TES investment; however, the resulting capacities overshoot the reference by $+41$–45 %, indicating that the affine correction overcompensates when the reference TES is moderate. This identifiability limitation of the affine model represents the main weakness of the approach; piecewise-linear could improve accuracy at the cost of additional variables.

**Peak diversity factor ($\alpha$).** The global peak diversity factor is $\alpha$=1.030, with per-cluster values ranging from 1.000 (Cluster 3, single building) to 1.043 (Cluster 4, 10 buildings with diverse thermal behaviors). These modest values reflect the tertiary nature of the case study; larger corrections should be expected for more mixed-use (literature reports 15–33 % for residential buildings).

## 4.3. Computational performance and sensitivity analysis

Aggregation reduces the number of decision variables by $10\times$ (from $5.19 \times 10^6$ to $519 \times 10^3$) and solve time from 181 s to 21 s (S1) or 54–60 s (S2), yielding a $3\times$–$9\times$ speedup. The $\beta$ correction adds $160 \times 10^3$ constraints (24 %) and increases solve time by $\sim 2.7\times$ relative to S1. The precalculation phase ($\sim$151 s for 7 grid points) dominates the total S2 time but is fully parallelizable and computed once for reuse across multi-scenario studies.
Table 7 reports the sensitivity to the three main parameters: precalculation grid resolution, number of spatial clusters $K$, and number of representative days $R_D$.

**Table 7**: Sensitivity analysis. (a) Grid resolution (random, $K$=5, $R_D$=12). (b) Spatial clusters ($R_D$=12, random). (c) Representative days ($K$=5, random). Bold: selected operating point.

**(a) Grid resolution**

| Pts | $\varepsilon^{S2}_{cost}$ [%] | Precalcul [s] | Solve [s] | Total [s] |
|---|---|---|---|---|
| 3 | −4.8 | 102 | 60 | 162 |
| 5 | −3.5 | 118 | 60 | 178 |
| **7** | **−3.3** | **151** | **60** | **211** |
| 10 | −3.0 | 237 | 65 | 302 |
| 15 | −2.9 | 469 | 67 | 536 |

**(b) Number of spatial clusters $K$**

| $K$ | $\varepsilon^{S1}_{cost}$ [%] | $\varepsilon^{S1}_{PV}$ [%] | $\varepsilon^{S2}_{cost}$ [%] | $\varepsilon^{S2}_{PV}$ [%] | Solve [s] |
|---|---|---|---|---|---|
| 1 | −32.5 | +15.2 | −16.6 | +15.2 | 7 |
| 3 | −29.6 | +15.2 | −11.2 | +6.6 | 36 |
| **5** | **−23.4** | **+15.2** | **−5.6** | **+8.9** | **60** |
| 10 | −12.2 | +9.9 | −1.9 | −3.0 | 117 |
| 15 | −7.3 | +3.7 | +0.2 | +3.3 | 148 |
| 20 | −6.8 | −0.5 | +0.6 | −2.1 | 191 |
| 25 | −5.8 | +0.5 | +1.4 | −1.0 | 223 |

**(c) Number of representative days $R_D$**

| $R_D$ | $\varepsilon^{S1}_{cost}$ [%] | $\varepsilon^{S1}_{PV}$ [%] | $\varepsilon^{S1}_{Bat}$ [%] | $\varepsilon^{S2}_{cost}$ [%] | $\varepsilon^{S2}_{PV}$ [%] | $\varepsilon^{S2}_{Bat}$ [%] | Solve [s] |
|---|---|---|---|---|---|---|---|
| 4 | −24.3 | +15.2 | −2.8 | −5.8 | +9.3 | −6.0 | 20 |
| 8 | −24.3 | +15.2 | −2.6 | −5.7 | +9.8 | −4.9 | 43 |
| **12** | **−23.4** | **+15.2** | **+0.4** | **−5.6** | **+8.9** | **−0.5** | **60** |
| 20 | −21.2 | +15.2 | +0.5 | −4.8 | +8.7 | +0.8 | 138 |
| 30 | −20.5 | +15.2 | +0.5 | −4.3 | +8.2 | −0.3 | 205 |

**The grid resolution sensitivity (panel a)** shows diminishing returns: cost error decreases from –4.8 % to –2.9 % as the grid is refined, with 1.5 pp gain from 3 to 7 points versus only 0.4 pp from 7 to 15. Solve time remains constant ($\sim$60–67 s), confirming that grid resolution does not affect MILP complexity. **The spatial sensitivity (panel b)** reveals that aggregation error is overwhelmingly spatial: S1 cost error decreases monotonically from –32.5 % ($K$=1) to –5.8 % ($K$=25), while S1 PV error saturates at 15.2 % for $K \leq 5$ and only decreases once finer clustering resolves per-building constraints. The $\beta/\alpha$ correction provides consistent improvement, reducing cost error by $2\times$ to $4\times$ at coarse resolutions ($K \leq 5$). At $K \geq 15$, the correction produces near-zero residual error ($+0.2$–1.4 %). **The temporal sensitivity (panel c)** confirms the spatial dominance: S1 PV error is constant at 15.2 % across all $R_D$, and S1 cost error remains stable (–20.5 % to –24.3 %). Storage accuracy is the exception, battery capacity error improves from $-2.8$% at $R_D$=4 to 0.4 % at $R_D$=12 as additional representative days capture charge–discharge pattern diversity. It also shows that the S2 cost

error varies by less than 1.5 pp between $R_D$=12 and $R_D$=30. Consequently, the reported errors $\varepsilon_{agg}$ and $\varepsilon_{pred}$ are overwhelmingly driven by spatial aggregation, not by the temporal resolution mismatch between scenarios.

The operating point $K$=5, $R_D$=12 offers the best compromise: $|\varepsilon^{S2}_{cost}|$=5.6 % with a total computation time of ∼211 s (151 s precalculation + 60 s solve), representing a 3× speedup over S0-T while reducing the prediction gap from −32 % to −18 %. The precalculation phase (151 s for 7 grid points) is parallelizable and, crucially, needs to be computed only once per clustering configuration: since $\beta$ and $\alpha$ depend solely on the techno-physical parameters that define the clustering, they remain valid across any variation of parameters that do not alter the clustering structure, such as energy prices or investment costs. This makes the mortized cost negligible in multi-scenario or multi-objective studies.

# 5. Conclusions

This paper proposed a preprocessing methodology correcting spatial aggregation errors in MILP optimization of multi-energy building portfolios. Three biases were identified and characterized: coverage error (self-consumption overestimation), pooling error (storage overestimation), and peak diversity error (heating capacity undersizing). The corrective coefficients $\beta$ and $\alpha$, computed through simulation and multivariate regression, are embedded as affine functions of decision variables directly into MILP constraints, preserving problem linearity.

Applied to a 50-building tertiary case study with $K$=5 spatial clusters and $R_D$=12 representative days over a 20-year horizon, the correction reduces the MILP cost error from −23.4 % to within −5.6 % to −7.5 % of the nodal reference, while maintaining a 3× solve time speedup. Sizing errors are substantially reduced for all subsystems except TES, whose correction is strongly strategy-dependent: the affine regression fails for uniform and proportional deployment strategies due to insufficient variance for slope estimation, while concentrated and random strategies achieve acceptable accuracy. This highlights that the effectiveness of the correction depends not only on the regression quality but also on how optimized capacities are redistributed across buildings.

A central finding is the distinction between optimization cost and deployment cost. The prediction gap $\varepsilon_{pred}$, measuring how reliably the aggregated model predicts the true system cost, is reduced by the correction from −31.7 % (S1) to −18.1 % (S2, uniform). This improvement is the primary practical benefit for investment planning: a decision-maker using S2 budgets within 18 % of reality, versus a 32 % shortfall with S1. The CAPEX and OPEX decomposition reveals that all deployed costs exceed S0-T due to OPEX penalties from imperfect intra-cluster capacity redistribution. When a cluster-level HP investment is disaggregated across individual buildings, under-endowed buildings incur high gas backup costs that outweigh the CAPEX savings of over-endowed ones, pushing total deployed costs above the nodal reference. The uncorrected model (S1) paradoxically achieves marginally lower deployed costs than S2 precisely because its systematic HP oversizing reduces the number of under-endowed buildings, limiting gas backup exposure at the expense of larger investment.

Future work should focus on two directions: (i) extending the framework to multi-objective formulations that maximize renewable energy penetration rather than minimizing cost alone; and (ii) validating the framework on larger and more heterogeneous portfolios where the aggregation biases, and therefore the correction benefits, are expected to be more pronounced.

# Acknowledgments

The authors gratefully acknowledge Akajoule for providing the ARX thermal demand model used to generate building-level heating profiles. The authors also thank Nantes Métropole for contextual data on the municipal building portfolio. This work is being carried out as part of the OSER project; a project carried out with financial support from ADEME.

# Nomenclature

$C$ installed capacity, kW or kWh
$D$ demand, kW
$G$ global horizontal irradiance, kW/m$^2$
$K$ number of spatial clusters
$N$ number of buildings
$P$ electrical power, kW
$Q$ thermal power, kW
$R_D$ number of representative days
$w_d$ weight of representative day $d$

**Greek symbols**

$\alpha$ peak diversity factor
$\beta$ corrective coefficient
$\gamma$ regression slope coefficient
$\varepsilon$ relative error, %

**Subscripts and superscripts**

$c$ cluster $c$
$n$ node $n$
$t$ time $t$
*ref* reference
*ch/dis* charge / discharge
*cov/pool* coverage / pooling
*prod/stor* production / storage
*sc* self-consumed

**Abbreviations**

*CSC* Collective self-consumption
*GB* Gas boiler
*HP* Heat pump
*ISC* Individual self-consumption
*MES* Multi-energy system
*MILP* Mixed-integer linear programming
*PV* Photovoltaic
*TES* Thermal energy storage

# References


[1] European Parliament and Council of the European Union. Directive (EU) 2024/1275 on the energy performance of buildings (recast). *Official Journal of the European Union*, L, 2024.

[2] European Commission. "Fit for 55": delivering the EU's 2030 Climate Target on the way to climate neutrality. COM(2021) 550 final, 2021.

[3] Mancarella P., MES (multi-energy systems): An overview of concepts and evaluation models. *Energy* 2014;65:1–17.

[4] Klemm C., Vennemann P., Modeling and optimization of multi-energy systems in mixed-use districts: A review. *Renew. Sustain. Energy Rev.* 2021;135:110206.

[5] Mancò G., Tesio U., Guelpa E., Verda V., A review on multi energy systems modelling and optimization. *Appl. Therm. Eng.* 2023;228:121871.

[6] Schütz T., Schiffer L., Harb H., Fuchs M., Müller D., Optimal design of energy conversion units and envelopes for residential building retrofits using a comprehensive MILP model. *Appl. Energy* 2017;185:1–15.

[7] Wirtz M., Kivilip L., Remmen P., Müller D., Design optimization of multi-energy systems using mixed-integer linear programming: Which model complexity and level of detail is sufficient? *Energy Convers. Manag.* 2021;240:114249.

[8] Kotzur L. et al., Impact of different time series aggregation methods on optimal energy system design. *Renew. Energy* 2018;117:474–487.

[9] Poncelet K. et al., Selecting representative days for generation expansion planning. *IEEE Trans. Power Syst.* 2017;32(3):1936–1948.

[10] Nahmmacher P. et al., Carpe diem: A novel approach to select representative days. *Energy* 2016;112:430–442.

[11] Hoffmann M. et al., A review on time series aggregation methods for energy system models. *Energies* 2020;13(3):641.

[12] Teichgraeber H., Brandt A.R., Time-series aggregation for the optimization of energy systems. *Renew. Sustain. Energy Rev.* 2022;157:111984.

[13] De Jaeger I. et al., A building clustering approach for urban energy simulations. *Energy Build.* 2020;208:109671.

[14] Yan Y. et al., Design and optimal siting of regional heat-gas-renewable energy system. *Energy Convers. Manag.* 2020;217:112963.

[15] Teichgraeber H., Brandt A.R., Clustering methods to find representative periods. *Appl. Energy* 2019;239:1283–1293.

[16] Schütz T. et al., Comparison of clustering algorithms for typical demand days. *Renew. Energy* 2018;129:570–582.

[17] Eggimann S. et al., Spatiotemporal upscaling errors of building stock clustering. *Energy Build.* 2022;258:111844.

[18] Johari F. et al., Urban building energy modeling from geo-referenced data. *Sustain. Cities Soc.* 2023;96:104664.

[19] Reinert C. et al., This is SpArta: Rigorous optimization of regionally resolved energy systems. *arXiv:2302.05222*, 2023.

[20] Patil S. et al., Advanced spatial and technological aggregation scheme. *Energies* 2022;15(24):9517.

[21] European Commission. Report on energy sharing. Publications Office of the European Union; 2025.

[22] Lowitzsch J., Hoicka C.E., van Tulder F.J., Renewable energy communities under the 2019 European Clean Energy Package. *Renew. Sustain. Energy Rev.* 2020;122:109489.

[23] Luthander R., Widén J., Nilsson D., Palm J., Photovoltaic self-consumption in buildings: A review. *Appl. Energy* 2015;142:80–94.

[24] Karalus S. et al., Analysing grid-level effects of photovoltaic self-consumption. *Energies* 2023;16(7):3059.

[25] Terlouw T. et al., Multi-objective optimization of energy arbitrage in community energy storage systems. *Appl. Energy* 2019;242:642–654.

[26] Weissmann C. et al., Analysis of heating load diversity in German residential districts. *Energy Build.* 2017;139:302–313.

[27] Wang Z. et al., Sizing of district heating systems based on smart meter data. *Energy* 2020;193:116780.

[28] Lloyd S.P., Least squares quantization in PCM. *IEEE Trans. Inform. Theory* 1982;28(2):129–137.

[29] Kaufman L., Rousseeuw P.J., *Finding Groups in Data*. New York, USA: Wiley; 1990.

[30] Arthur D., Vassilvitskii S., k-means++: The advantages of careful seeding. In: *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*; 2007 Jan 7–9; New Orleans, USA. pp. 1027–1035.

[31] Park H.-S., Jun C.-H., A simple and fast algorithm for K-medoids clustering. *Expert Syst. Appl.* 2009;36(2):3336–3341.