Energy Conversion and Management 358 (2026) 121481

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Estimating residential flexibility from smart meter data: A bilevel framework for demand response aggregation

Reda El Makroum [a],*, Sebastian Zwickl-Bernhard [a,b], Lukas Kranzl [a], Hans Auer [a,b]

[a] Energy Economics Group (EEG), Technische Universität Wien, Gußhausstraße 25-29/E370-3, 1040 Vienna, Austria
[b] Department of Industrial Economics and Technology Management, The Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:
Demand response
Bilevel optimization
Flexibility estimation
Smart meter data
Residential aggregation
Home energy management systems

ABSTRACT

Existing bilevel optimization frameworks for residential demand response assume known household flexibility characteristics, limiting their applicability to real-world portfolios where flexibility is estimated from consumption data. This paper develops a data-driven bilevel framework that addresses both challenges: estimating flexibility from smart meter data and optimizing aggregator-household coordination under behavioral uncertainty. The estimation component applies an appliance-agnostic decomposition to 15-minute resolution data, extracting flexible loads without requiring device-level metering. These estimates feed into a Stackelberg game where the aggregator optimizes flexpoint-based rewards and households respond by scheduling flexible consumption. The framework captures behavioral heterogeneity by distinguishing automated households providing deterministic responses from manual households exhibiting probabilistic participation with willingness decay. Simulation across 12 German households over 364 days enables 20.7 MWh of shifted load, generating 1029 EUR in cost savings. For this portfolio, participation frequency drove economic performance more strongly than event magnitude, with automated households contributing the majority of total value despite comprising half the portfolio. The framework provides aggregators with a scalable approach to quantify and coordinate distributed residential flexibility.

1. Introduction

The global energy transition requires all sectors to fundamentally rethink how energy is consumed and how to achieve deep decarbonization across the entire energy system. This transformation is particularly critical in the residential sector, which accounts for 27% of global electricity consumption [1] and represents a significant untapped resource for system flexibility [2]. As energy systems evolve toward high renewable energy penetration, demand-side flexibility becomes essential for balancing variable generation and ensuring grid stability [3]. The residential sector, with its diverse consumption patterns, distributed energy resources, and growing smart technology adoption [4], emerges as a cornerstone for achieving both decarbonization objectives and system flexibility goals through coordinated demand response (DR) initiatives.

DR programs provide the coordination mechanism for this distributed flexibility through incentive-based schemes that offer households financial rewards for shifting consumption away from peak periods [5]. Effective DR program design requires frameworks that quantify household flexibility potential while accounting for different types of households and their corresponding participation behaviors. The objective of this paper is to develop such a framework, where we model aggregator-household coordination as a hierarchical optimization problem with strategic interactions between actors.

Unlike dispatchable generation, residential flexibility is challenging to measure directly. While appliance-level submetering would provide precise flexibility estimates, such instrumentation is rarely deployed at scale. Non-intrusive load monitoring (NILM) offers an alternative but typically requires high-frequency sampling rates to detect appliance signatures [6], whereas widely deployed smart meters commonly record consumption at 15-minute intervals [7]. This necessitates methods that extract flexibility potential directly from low-resolution aggregate data. Beyond data resolution, effective DR programs must be inclusive, accommodating households regardless of whether they have automated energy management systems or rely on manual participation [8]. Even with known device portfolios, aggregators must model how their incentives and household responses interact to translate technical capacity into realized flexibility [9]. The remainder of this paper is structured as follows. Section 2 reviews flexibility estimation methods, portfolio aggregation approaches, and bilevel optimization frameworks. Section 3 presents the methodology, including smart meter decomposition,

* Corresponding author.
E-mail addresses: reda.el.makroum@tuwien.ac.at, elmakroum@eeg.tuwien.ac.at (R. El Makroum).

https://doi.org/10.1016/j.enconman.2026.121481
Received 2 January 2026; Received in revised form 27 March 2026; Accepted 9 April 2026
Available online 21 April 2026

bilevel optimization framework, and behavioral modeling. Section 4 presents computational results. Section 5 discusses implications for DR program design. Section 6 concludes with key findings and future research directions.

2. Literature review

Research on residential flexibility estimation has grown rapidly, driven by renewable energy integration, smart grid deployment, and the critical role of demand-side resources in grid stability. However, significant implementation barriers persist. Parrish et al. [10] systematically review residential DR trials and identify barriers to consumer engagement including perceived risk and control, complexity and effort, and consumer characteristics, noting considerable variation in user engagement across programs. Despite these barriers, system-level benefits justify continued research efforts. Miri and McPherson [11] demonstrate DR programs deliver flexibility at lower cost than supply-side solutions, achieving 7.7% avoided wind curtailment and up to 15% household bill savings in Alberta's zero-emission power system scenarios. Thrän et al. [12] introduce a levelised cost framework for comparing DR schemes against storage technologies. Their analysis shows that heat pump-based DR and electric vehicle (EV) smart charging can be cost-competitive with conventional storage for certain applications.

2.1. Flexibility estimation using machine learning

Advanced data-driven approaches leveraging machine learning have gained traction for flexibility estimation. Recurrent neural networks have proven particularly effective. Papias et al. [13] employ Long Short-Term Memory (LSTM) networks for next-day load variation forecasting to construct flexibility baseline loads using real-world prosumer data. Their subsequent work [14] develops a BiLSTM-based methodology achieving R^2 values exceeding 0.85 for consumption forecasting and 0.97 for photovoltaic (PV) production across diverse building typologies in Austria and Ireland. Zhang et al. [15] propose a hybrid CNN-GRU architecture for building flexibility potential assessment, incorporating consumer psychology factors into the prediction model. Addressing prediction uncertainty, Bampoulas et al. [16] develop a Bayesian deep-learning framework for assessing residential flexibility with multicomponent energy systems, achieving uncertainty-aware predictions that capture the stochastic nature of household behavior. Haapaniemi et al. [17] take an interpretability-focused approach, applying machine learning and Shapley values to quantify implicit price flexibility using data from over 6000 customers.

While these data-driven methods demonstrate strong predictive performance, they primarily focus on forecasting aggregate flexibility without addressing the coordination mechanisms required for translating estimated flexibility into actionable DR programs.

2.2. Portfolio aggregation and consumer behavior modeling

Large-scale aggregation and clustering methodologies have been developed to handle diverse residential portfolios [18]. Petrucci et al. [19] apply time-series k-means clustering to 2031 thermal load profiles using smart thermostat data and gray-box models, achieving a 94% increase in daily load factor through coordinated DR. Kallel et al. [20] advance interpretability through Gaussian mixture models with feature importance analysis. Beyond consumption data alone, Vahedi et al. [21] demonstrate that incorporating sociological and meteorological factors improves residential peak demand classification. Taking a behavioral perspective, Wang et al. [22] simulate interrelated load profiles by extracting activities from time use surveys, while Zhang et al. [23] propose a cloud–edge collaboration framework incorporating consumer psychology through fuzzy response mechanisms.

While these approaches demonstrate effective portfolio management and behavioral modeling, they treat households as passive assets rather than strategic actors, generally assuming exogenous response patterns rather than modeling the strategic interaction between aggregator incentives and household decisions.

2.3. Bilevel optimization for demand response coordination

Bilevel optimization models hierarchical decision-making where an upper-level leader anticipates the optimal response of lower-level followers [24]. This Stackelberg game structure captures strategic interactions between parties with potentially conflicting objectives [25], making it well-suited for aggregator-household coordination where incentive design must account for rational consumer responses.

Zhang et al. [26] propose a bilevel method for flexibility feature extraction and user clustering using smart meter data from over 5000 residential EV users in China, applying Pearson correlation and Gaussian Mixture Models to identify six charging segment categories and five user categories for targeted smart charging strategies. Wang et al. [27] present a score-based incentive DR model combining construal level theory and fuzzy theory to model consumers' power-score redeem behavior, optimizing incentive parameters at the upper level while incorporating comfort and redeem behavior at the lower level. Lin et al. [28] extend this psychological focus through a trilayer Stackelberg game capturing economic, environmental, and behavioral dimensions of incentive-based DR. Barala et al. [29] develop a bilevel framework for coordination between Distribution System Operators and Virtual Energy Storage Systems using thermostatically controlled loads, incorporating building dynamics and occupant activities for flexibility estimation with dynamic pricing mechanisms. Yin et al. [9] formulate a distributionally robust bilevel model for distribution networks under uncertain renewables and temperatures, providing tractable mixed-integer linear programming (MILP) reformulation with out-of-sample guarantee while demonstrating pricing strategies that address equity considerations among heterogeneous households. Dong et al. [30] introduce a bilevel framework for multi-park integrated energy systems with categorized DR models for industrial, residential, and commercial parks, utilizing unified adaptive robust optimization to address renewable uncertainties. Davoudi et al. [31] model a residential complex as an energy hub using a single-leader multi-follower Stackelberg game, formulating interactions as bilevel optimization with upper-level energy management and lower-level multi-period optimal power and thermal flows.

Most closely related to this work, Rasouli et al. [32] use genetic algorithms at the upper-level combined with MILP solvers at the lower-level to determine optimal reward signals for flexibility aggregation and ancillary services market participation. Their approach assumes known flexibility characteristics across 20 synthetic prosumer profiles with predetermined resource parameters, differing fundamentally from our work which estimates household flexibility from real smart meter data without prior knowledge of appliance configurations.

While these bilevel frameworks successfully model strategic interactions and consumer behavior, they predominantly assume complete knowledge of household flexibility characteristics and deterministic participation patterns. Existing work does not adequately address the challenge of estimating flexibility from limited data nor incorporate realistic behavioral uncertainty. Furthermore, practical budget allocation mechanisms that adapt to actual household responses during sequential decision-making remain unexplored in current bilevel DR literature. Table 1 summarizes the key elements of recent bilevel DR studies and highlights the specific contributions of the present work.

To address these gaps, we formulate the following research question: How can aggregators estimate residential flexibility from smart meter data and determine optimal reward signals that account for behavioral heterogeneity across automated and manual participation modes within a bilevel optimization framework?

2.4. Key contributions

Building on the identified gaps in existing literature, this paper makes the following contributions:

Table 1
Summary of recent bilevel optimization frameworks for residential demand response.

Reference	Approach	Flexibility input	Optimization method	Key distinction
Wang (2024)	Score-based incentive DR	Manual preference input; fuzzy theory	Bilevel with construal level theory	Fuzzy consumer psychology; no flexibility estimation
Lin (2022)	Trilayer Stackelberg DR	Parameterized psychological factors	Trilayer Stackelberg game	Economic-environmental-behavioral modeling; assumes known flexibility
Zhang (2025a)	Cloud-edge collaboration	Fuzzy response mechanisms	Multi-objective bilevel	Consumer psychology modeling; no smart meter decomposition
Zhang (2025b)	EV user clustering	Smart meter data from 5000 EVs	Bilevel with Gaussian mixture clustering	EV-specific; assumes known charging parameters
Yin (2025)	Distributionally robust pricing	Uncertain renewables and temperature	MILP with distributionally robust optimization	Equity-focused pricing; deterministic household responses
Barala (2024)	Coordination with virtual energy storage	RC model with building dynamics	Bilevel with dynamic pricing	Thermostatically controlled loads; predetermined flexibility
Dong (2025)	Multi-park integrated energy system	Categorized DR (industrial, residential, commercial)	Adaptive robust optimization	Park-level aggregation; no behavioral heterogeneity
Davoudi (2024)	Stackelberg energy hub	Multi-follower negotiation	Bilevel with multi-period optimal power flow	Energy hub optimization; known resource parameters
Rasouli (2025)	Flexibility aggregation for ancillary services	20 synthetic prosumer profiles	Genetic algorithm + MILP	Closest to present work; assumes predetermined flexibility parameters
Present study	**Flexpoint-based Stackelberg DR**	**Appliance-agnostic decomposition from 15-minute smart meter data**	**Bilevel framework with sequential household processing**	**Data-driven appliance-agnostic flexibility estimation; behavioral heterogeneity**

- **Data-driven bilevel framework from real smart meter data.** In contrast to existing bilevel DR frameworks that assume known flexibility characteristics through synthetic prosumer profiles or predetermined parameters, we develop a data-driven bilevel optimization framework that estimates household flexibility directly from real smart meter consumption data using variability patterns that capture flexibility potential.
- **Appliance-agnostic decomposition at 15-minute resolution.** We introduce an appliance-agnostic decomposition method that extracts flexibility potential from 15-minute resolution smart meter data using percentile bounds and standard deviation indicators, and integrate these estimates into a bilevel optimization framework that captures aggregator-household strategic interactions, operating with estimated consumption parameters. This preserves household data privacy, as the method requires no appliance-level data or preference information.
- **Behavioral heterogeneity across automated and manual modes.** We incorporate behavioral heterogeneity by distinguishing between automated energy management systems and manual participation modes within the bilevel framework. Unlike existing bilevel DR frameworks that assume deterministic homogeneous responses, this captures realistic behavioral uncertainty across participation modes, enabling aggregators to design tailored reward strategies for diverse household portfolios.

3. Methodology

Fig. 1 presents a schematic overview of the proposed framework. The methodology proceeds in two stages. In the first stage, described in Section 3.1, historical smart meter data and day-ahead electricity prices are used to estimate household flexibility through statistical load decomposition. In the second stage (Sections 3.2 and 3.3), these flexibility estimates feed into a bilevel optimization framework where the aggregator sequentially coordinates household responses to evaluate the DR value capturable across the portfolio.

3.1. Flexibility estimation from smart meter data

To estimate household flexibility from 15-minute smart meter data, we adapt and extend the methodology developed by Shi and Jiao [33], which quantifies demand response capacity without requiring appliance-level metering. The original framework classifies households into flexibility categories using machine learning; we extend this to perform load decomposition, extracting the specific flexible energy chunks and their temporal characteristics needed for bilevel optimization. We apply the framework to 15-minute resolution smart meter data from 12 German households across 364 days of 2023, combined with day-ahead electricity prices from the ENTSO-E Transparency Platform [34]. This retrospective analysis quantifies flexibility that existed in historical consumption patterns, demonstrating achievable DR value under known conditions. Since the method operates exclusively on aggregate smart meter data, no appliance configurations, household preferences, or comfort constraints are required as input, preserving household data privacy by design.

The approach is grounded in the principle that household flexibility can be reliably estimated from observable consumption patterns and behavioral characteristics present in smart meter data [35]. The fundamental insight is that households exhibiting natural variability in their consumption patterns possess latent flexibility that can be activated through demand response signals [36]. This behavioral approach captures real-world comfort constraints, lifestyle limitations, and aggregated flexibility from all household sources while reflecting actual willingness and ability to respond to grid signals.

The methodology establishes a Typical Electricity Consumption Profile (TECP) for each household using median consumption values across different time periods under normal pricing conditions. This baseline serves as the reference point for all flexibility calculations and requires a minimum of three months of smart meter data for stable characterization. From this baseline, the approach develops key flexibility indicators that capture different aspects of household responsiveness. Daily consumption sum reflects household size and lifestyle, Nomenclature

Sets			
$\mathcal{H}$	Set of households	$\mathcal{T}$	Set of time slots
Indices			
d	Day index	h	Household index
i	Processing order index	t	Time slot index
Parameters			
α	Reward regularization coefficient	B_0	Initial daily budget
$\beta_{h,t}$	Median baseline consumption (kW)	e_h	Energy per slot (kWh)
γ	Confidence adjustment factor	L_h	Flexible load duration (slots)
δ	Willingness decay rate	M	Big-M parameter
η	Budget adjustment rate	P_t	Electricity price (EUR/kWh)
ϕ	Utilization threshold	$Q^{max}_{base,t}$	Upper percentile bound (kW)
σ_R	Participation sensitivity	$Q^{min}_{base,t}$	Lower percentile bound (kW)
σ_t	Standard deviation at slot t	T	Total time slots per day
θ	Participation threshold		
Variables			
$\Gamma_{h,d}$	Participation probability	r_t	Flexpoint reward per kWh at slot t
B_{agg}	Available aggregator budget	$R_{h,d}$	Expected rewards
B^h_{agg}	Remaining budget for household h	$R^{allocated}_i$	Allocated flexpoints
B_d	Budget on day d	U_d	Budget utilization rate
B_{total}	Total flexpoint budget	$u_{h,t}$	Rising edge binary variable
$C^{flex}_{h,d,t}$	Flexible consumption (kW)	$w_{h,d}$	Willingness factor
E^{flex}_h	Total flexible energy (kWh)	$y_{h,t}$	Binary activation variable
$I_{h,d}$	Participation indicator (binary)	$y^*_h(t)$	Optimal activation decision
λ_h	Duration constraint dual variable	$z_{h,t}$	Complementarity binary variable
$\mu_{h,t}$	Upper bound dual variable		
Subscripts and Superscripts			
agg	Aggregator	max	Maximum bound
$base$	Baseline profile	min	Minimum bound
$flex$	Flexible component	$total$	Total (portfolio-level)
Abbreviations			
DR	Demand Response	KKT	Karush–Kuhn–Tucker
EV	Electric Vehicle	MILP	Mixed-Integer Linear Program
HEMS	Home Energy Management System	NILM	Non-Intrusive Load Monitoring
IoT	Internet of Things	TECP	Typical Electricity Consumption Profile

while period-specific consumption reveals temporal availability patterns. Standard deviation by time period measures natural consumption volatility, peak–valley ratio indicates load shifting potential, and maximum and minimum flexibility bounds define response capacity limits.

The implementation in this study extends the methodology to decompose household load profiles into flexible and inflexible components based on the estimated flexibility indicators. The decomposition algorithm operates at 15-minute time slot resolution, utilizing the 25th and 75th percentile bounds (Q^{min}_{base} and Q^{max}_{base}) and standard deviation (σ_t) to establish confidence-weighted baselines for each time slot. The confidence weighting adjusts the median baseline using the standard deviation indicator. The total daily flexible load is calculated as:

$$\sum_{t=1}^{96} C^{flex}_{h,d,t} = \sum_{t=1}^{96} \max(0, C_{h,d,t} - (\beta_{h,t} - \gamma \cdot \sigma_t \cdot (\beta_{h,t} - Q^{min}_{base,t})))$$

where $C^{flex}_{h,d,t}$ is the flexible consumption for household h on day d at time slot t, $C_{h,d,t}$ is actual consumption, $\beta_{h,t}$ is the baseline consumption profile, σ_t is the standard deviation at time slot t, $Q^{min}_{base,t}$ is the lower percentile bound, γ is the confidence adjustment factor (0.20), and the confidence-weighted baseline represents the essential non-flexible load. The algorithm applies a minimum threshold filter of 0.2 kW to remove trivial flexible loads, as threshold-based filtering improves estimation accuracy by excluding small fluctuations that may represent disturbances rather than genuine schedulable events [37]. Finally, only the largest continuous flexible load period per day is retained for aggregation. This conservative modeling choice reflects practical DR program design where single coordinated events simplify both aggregator scheduling and household participation, though it may underestimate total flexibility potential for households capable of multiple daily response events. Because the decomposition operates on aggregate consumption, it does not distinguish between load types, and these filtering steps naturally retain sustained flexible loads over short consumption events. In practice, the flexible components are most likely to correspond to deferrable loads such as washing machines, dishwashers, or EV charging, as these produce sustained consumption periods consistent with the continuous block retained by the method. Fig. 2 illustrates this process for a representative household.

This methodology is applied to smart meter data from 12 households in Germany, spanning the entire year of 2023. The decomposition is illustrated using data from one of these households during the first week of January 2023. Panel (a) shows raw 15-minute smart meter data. Panel (b) demonstrates decomposition into flexible (green) and non-flexible (blue) components based on TECP analysis. Panel (c) shows the refined approach where only the largest continuous flexible load period per day is retained for aggregation.

This element is implemented in the optimization through constraint (7), which ensures at most one rising edge transition (from inactive to active flexibility provision), thereby guaranteeing that each household can contribute only a single continuous flexibility period per day. In this work, households are categorized into two distinct participation types. The first category comprises households equipped with a Home Energy Management System (HEMS), which can automatically respond to aggregator signals with ideal reliability. The second category represents manual participation households, where enthusiastic residents receive demand response signals through mobile app notifications and make discretionary participation decisions.

To ensure unbiased comparison between participation types, households were sorted by descending annual consumption and alternately assigned to HEMS (H1–H6) and manual (H7–H12) categories. This

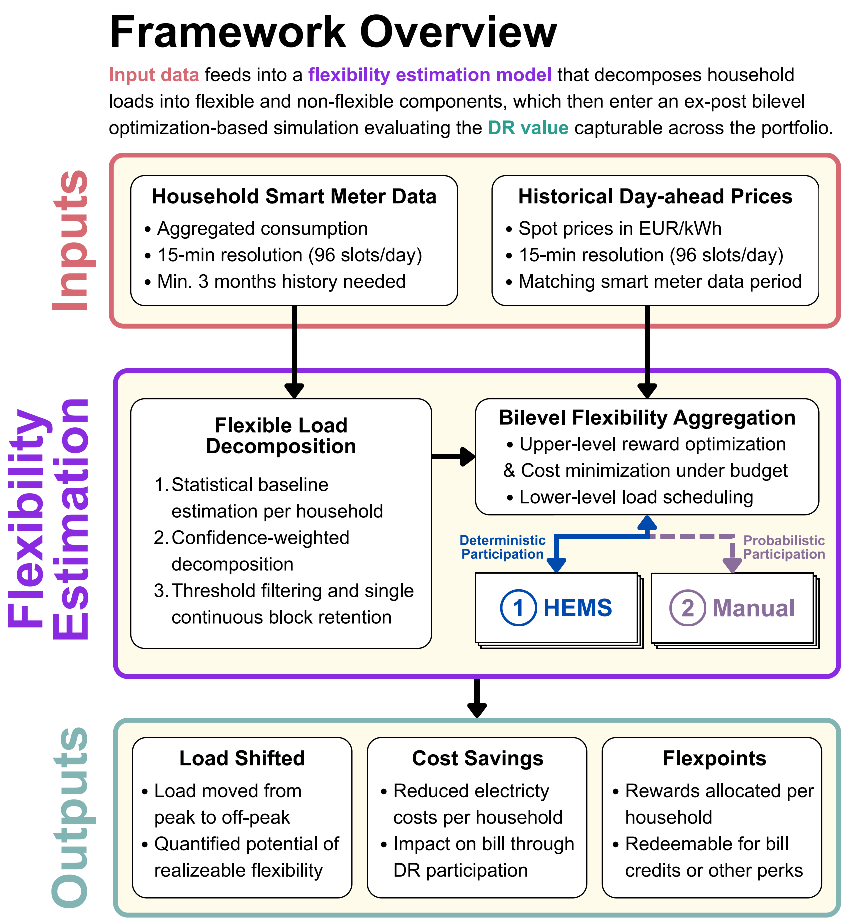

Fig. 1. Framework overview.

systematic assignment prevents consumption-based bias while maintaining comparable household sizes across both participation types. Fig. 3 presents the annual flexibility analysis results.

The stacked bar chart in Fig. 3 illustrates three distinct load components: non-flexible baseline load (blue), fragmented flexible load (light blue), and single continuous flexible load that can potentially be effectively aggregated (green). Households with diagonal hatching patterns represent manual participation, while solid patterns indicate HEMS-enabled households. With this flexibility decomposition complete (numerical results in Appendix A), the critical challenge becomes translating individual household patterns into reliable demand response program estimates. This requires modeling the strategic interactions between aggregators seeking profit maximization and households optimizing their utility under different participation modes.

3.2. Problem overview

This work models strategic interactions between a profit-maximizing aggregator and utility-maximizing households, where optimal reward allocations incentivize demand response participation while household consumption schedules respond to these incentives. Rewards are structured as flexpoints, a points-based system separate from electricity billing that enables aggregators to compensate households for flexibility provision through bill credits, vouchers, or charity donations. Flexpoints serve as an intermediary unit between the aggregator and households: the aggregator allocates flexpoints per kWh of shifted consumption (r_t), and their monetary value is determined by the aggregator based on revenue captured from flexibility market participation. The complexity stems from three key factors. First, limited budgets must be allocated across households with varying flexibility and participation modes. Second, HEMS households provide deterministic responses while manual households demonstrate probabilistic participation. Third, sequential processing requires committing budget allocations before observing aggregate portfolio outcomes. The physical realization of flexibility differs by household type: HEMS households automatically adjust the scheduling of IoT-enabled loads in response to aggregator signals, while manual households shift consumption themselves following app notifications. The Algorithm summarizes the complete framework.

This coordination challenge is formulated as a series of interconnected bilevel optimization problems, where each household's optimization constitutes an individual bilevel problem solved sequentially within the aggregator's budget allocation framework. The upper level represents the aggregator's strategic reward allocation decisions, while the lower level captures each household's optimal consumption scheduling response to the proposed rewards. The mathematical formulation addresses both deterministic HEMS households and stochastic manual households through differentiated modeling approaches. For HEMS households, the lower level problem represents a deterministic optimization where consumption scheduling follows predictable patterns based on technical constraints and reward signals. For manual households, the framework incorporates probabilistic participation decisions influenced by willingness factors, social dynamics, and fairness considerations, reflecting the behavioral uncertainties inherent in human-mediated demand response participation.

3.3. Bilevel optimization framework

3.3.1. Upper level problem

The aggregator's optimization problem represents the strategic decision-making layer where reward allocations are determined to

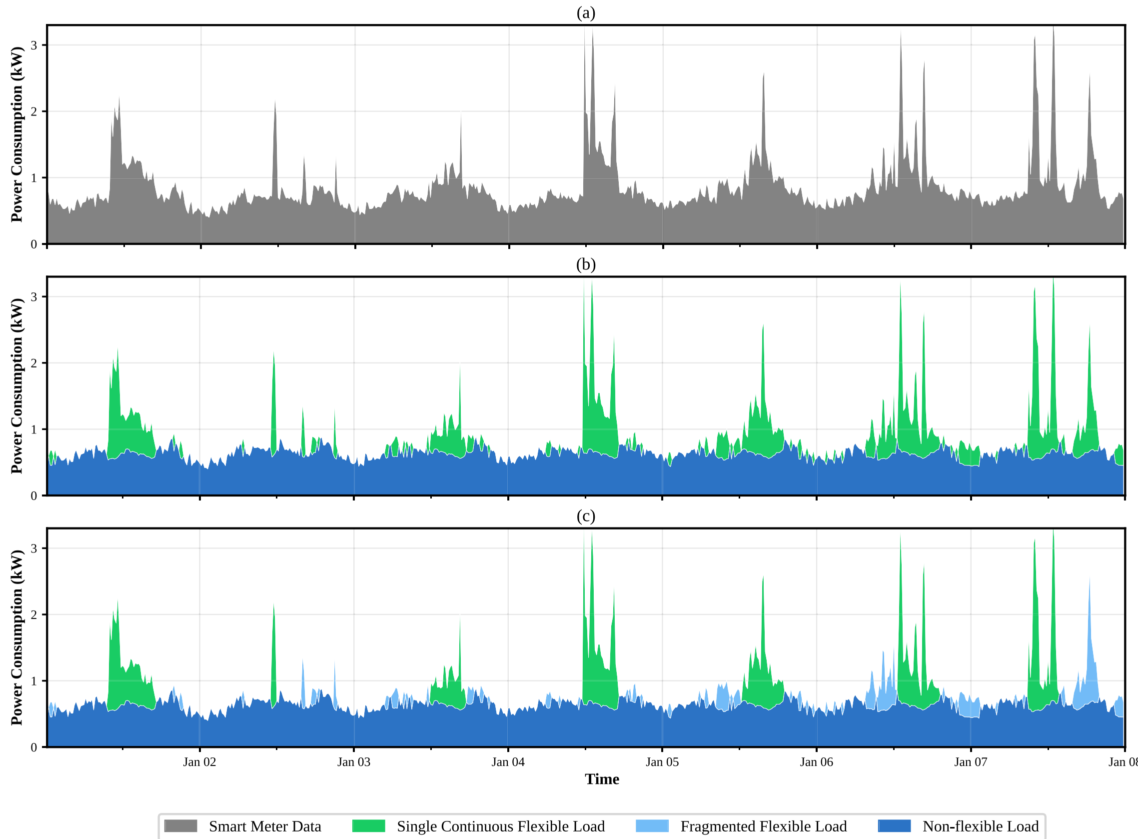

Fig. 2. Household Load Profile Decomposition into Flexible and Non-flexible Components Using Smart Meter Data Analysis. Blue: estimated non-flexible baseline; light blue: estimated fragmented flexible load; green: estimated single continuous flexible load retained for aggregation. Categories represent statistical decomposition from aggregate consumption.

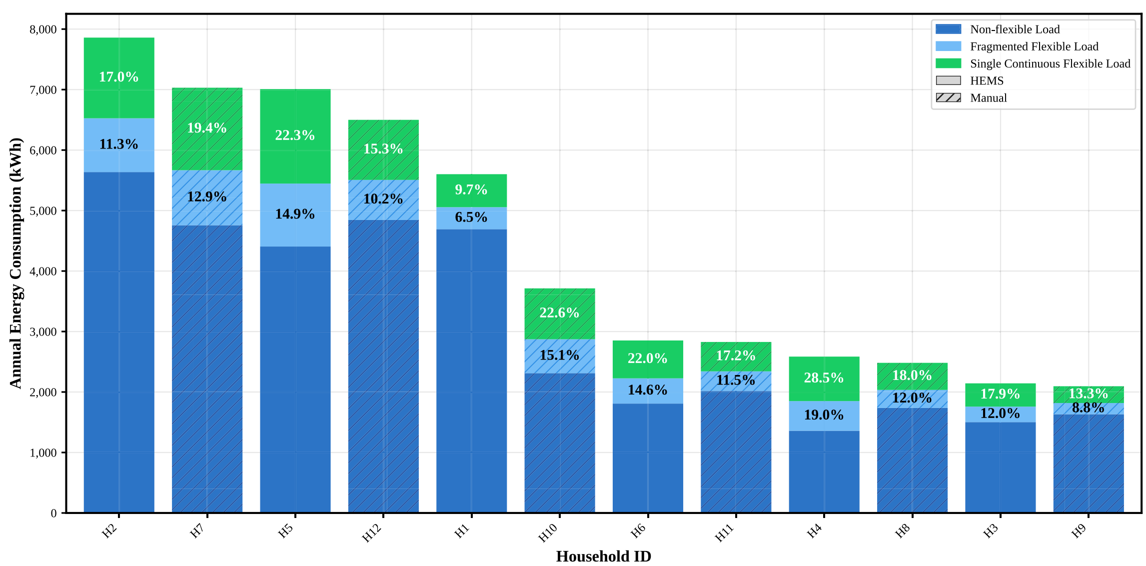

Fig. 3. Annual Load Profile Analysis: Fragmented vs Single Continuous Flexibility Across 12 Households in Schönbrunn, Germany (2023).

minimize total electricity costs while incentivizing optimal household participation. For each individual household $h \in \mathcal{H}$, the aggregator solves the following upper level optimization problem:

$$\min_{r_t} \quad \sum_{t \in \mathcal{T}} P_t \cdot y_h^*(t) \cdot e_h + \alpha \sum_{t \in \mathcal{T}} r_t^2 \tag{1}$$

subject to:

$$\sum_{t \in \mathcal{T}} r_t \cdot y_h^*(t) \cdot e_h \leq B_{agg} \tag{2}$$

$$r_t \geq 0, \qquad \forall t \in \mathcal{T} \tag{3}$$

The objective function minimizes electricity costs for the flexible load placement, where P_t represents dynamic 15-minute day-ahead

Algorithm Flexibility estimation framework.

Require: Smart meter data $C_{h,d,t}$ for households $h \in \mathcal{H}$.
Require: Day-ahead electricity prices P_t for D days.
Require: Initial daily budget B_0 and behavioral parameters θ, σ_R, δ.
Estimate flexibility E_h^{flex}, L_h for each household via TECP decomposition.
Categorize households into HEMS (H1–H6) and manual (H7–H12).
for $d = 1$ to D **do**
  Set daily budget B_d via adaptive adjustment (Eq. (16)).
  for each HEMS household h in randomized order **do**
    Solve bilevel optimization (Eqs. (1)–(9)).
    Deduct allocated rewards from B_d.
    if $B_d < B_{min}$ **then** terminate sequential processing.
    end if
  end for
  for each manual household h in randomized order **do**
    Evaluate participation probability $\Gamma_{h,d}$ (Eq. (10)).
    if household participates **then**
      Solve bilevel optimization (Eqs. (1)–(9)).
      Deduct allocated rewards from B_d.
      if $B_d < B_{min}$ **then** terminate sequential processing.
      end if
    end if
    Update willingness factor $w_{h,d+1}$ (Eq. (11)).
  end for
end for
Output: Shifted energy, cost savings, and flexpoints per household.

electricity prices, $y_h^*(t)$ denotes the household's optimal activation decision from the lower level, e_h is the energy per slot, r_t are flexpoints offered per kWh, and α is a regularization coefficient preventing excessive reward concentration in single time slots. Constraint (2) limits total reward expenditure to available allocation B_{agg}, representing remaining flexpoints after previous households' optimization. Constraint (3) ensures non-negative rewards.

3.3.2. Lower level problem

The household responds to the aggregator's reward signals by solving an optimization problem that maximizes utility from demand response participation while respecting operational constraints. For household h, the lower level problem is formulated as:

Building on the flexibility decomposition from Section 3.1, the household's optimization determines where to place the identified flexible energy chunk E_h^{flex} (total flexible energy in kWh) with duration L_h (number of time slots). The energy per slot when active is $e_h = E_h^{flex}/L_h$.

$$\max_{y_{h,t}} \sum_{t\in\mathcal{T}} r_t \cdot y_{h,t} \cdot e_h \tag{4}$$

subject to:

$$\sum_{t\in\mathcal{T}} y_{h,t} = L_h \tag{5}$$

$$u_{h,t} \geq y_{h,t} - y_{h,t-1}, \qquad \forall t \in \{2,\dots,T\} \tag{6}$$

$$\sum_{t=2}^{T} u_{h,t} \leq 1 \tag{7}$$

$$y_{h,t} \in \{0,1\}, \qquad \forall t \in \mathcal{T} \tag{8}$$

$$u_{h,t} \in \{0,1\}, \qquad \forall t \in \{2,\dots,T\} \tag{9}$$

The objective function maximizes rewards earned from flexibility activation, where $y_{h,t} = 1$ indicates the flexible load is active at time slot t and e_h is the constant energy consumption per slot. Constraint (5) ensures the flexible chunk occupies exactly L_h time slots, where L_h is the duration of the largest continuous flexible period identified in Section 3.1; since $e_h = E_h^{flex}/L_h$, this implicitly enforces energy conservation. Constraints (6)–(7) enforce contiguity through linearized rising edge detection: auxiliary binary variables $u_{h,t}$ capture transitions from inactive to active states, and limiting their sum to one ensures at most one contiguous activation period per day. This formulation determines the optimal temporal placement of the flexible energy chunk, shifting it from its baseline position to periods with higher rewards.

3.3.3. Behavioral modeling for manual households

While HEMS households provide deterministic responses to aggregator signals, manual households introduce stochastic participation patterns that require explicit behavioral modeling. Manual households receive demand response notifications but make discretionary participation decisions influenced by individual willingness. Although participation probability could be modeled using simple threshold-based rules or linear functions, a sigmoid function is utilized in this model to enable smooth probabilistic transitions and capture the diminishing sensitivity to reward differences that characterizes human decision-making:

$$\Gamma_{h,d} = w_{h,d} \cdot \sigma\left(\frac{R_{h,d} - \theta}{\sigma_R}\right) \tag{10}$$

where $w_{h,d}$ represents the willingness factor, σ is the sigmoid function, $R_{h,d}$ denotes expected rewards for household h on day d, θ is the participation threshold parameter representing the minimum reward level where participation becomes attractive, and σ_R controls the sensitivity around the threshold.

The willingness factor evolves dynamically based on participation decisions through a decay mechanism:

$$w_{h,d+1} = w_{h,d} \cdot (1 - \delta \cdot I_{h,d}) \tag{11}$$

where δ is the decay rate parameter and $I_{h,d}$ is a binary indicator equal to 1 if household h participated on day d. This decay reflects participation fatigue and the diminishing enthusiasm for repeated demand response events.

3.3.4. Single-level MILP formulation

The bilevel structure is addressed through a hybrid MILP formulation. Karush–Kuhn–Tucker (KKT) stationarity conditions for the household's reward-maximization problem create a threshold mechanism that identifies optimal activation slots, while contiguity constraints are enforced directly in the MILP rather than through dual variables, as this is computationally more efficient.

For the reward-maximization aspect, the Lagrangian introduces dual variable λ_h for the duration constraint and $\mu_{h,t} \geq 0$ for the implicit upper bound $y_{h,t} \leq 1$. The KKT stationarity condition yields:

$$r_t \cdot e_h + \lambda_h - \mu_{h,t} = 0, \quad \forall t \in \mathcal{T} \tag{12}$$

This creates a threshold mechanism where $-\lambda_h$ acts as the activation threshold: slots with $r_t \cdot e_h > -\lambda_h$ satisfy $\mu_{h,t} > 0$, which by complementary slackness implies $y_{h,t} = 1$. The dual variable λ_h adjusts such that exactly L_h slots exceed the threshold.

Complementary slackness for the upper bound constraint $y_{h,t} \leq 1$ is enforced using Big-M formulations:

$$\mu_{h,t} \leq M \cdot z_{h,t}, \qquad \forall t \in \mathcal{T} \tag{13}$$

$$1 - y_{h,t} \leq M \cdot (1 - z_{h,t}), \qquad \forall t \in \mathcal{T} \tag{14}$$

where $z_{h,t}$ are auxiliary binary variables and M is set sufficiently large to not constrain feasible solutions (10× the maximum reward rate in this implementation). These constraints enforce $\mu_{h,t} \cdot (1 - y_{h,t}) = 0$: either $\mu_{h,t} = 0$ (when $z_{h,t} = 0$) or $y_{h,t} = 1$ (when $z_{h,t} = 1$). The contiguity constraints (6)–(7) are included directly, and the aggregator's control over reward signals r_t shapes household responses toward contiguous activation windows. The complete formulation is solved as a single MILP using Gurobi.

Table 2
Model input parameters.

Parameter	Value
Initial daily budget (B_0)	10,000 flexpoints
Utilization threshold (ϕ)	90%
Adjustment rate (η)	±10%
Lower budget bound	$0.5 \times B_0$
Upper budget bound	$10 \times B_0$
Participation threshold (θ)	25–145 flexpoints
Participation sensitivity (σ_R)	10 flexpoints
Willingness decay rate (δ)	1%
Confidence adjustment factor (γ)	0.20
Minimum flexibility threshold	0.2 kW

3.3.5. Sequential processing

The bilevel optimization framework is solved sequentially for each household following a structured processing order. HEMS households are prioritized and processed first, followed by manual households. This prioritization reflects practical aggregator preferences for guaranteed participation while creating incentives for manual households to adopt an HEMS in the future. Within each category, household order is randomized to ensure equitable access to budget allocation. For each household h, the available budget is updated as:

$$B_{agg}^{h} = B_{total} - \sum_{i<h} R_i^{allocated} \tag{15}$$

where $R_i^{allocated}$ represents rewards allocated to previously processed households. HEMS households solve the deterministic bilevel optimization directly, while manual households first evaluate participation probability using Eq. (10) before proceeding with optimization.

Two operational mechanisms govern total reward budget allocation. First, the sequential processing terminates when available budget falls below the minimum viable allocation threshold. Second, daily budget allocation adapts dynamically based on previous day interactions through the following adjustment logic:

$$B_{d+1} = \begin{cases} B_d \cdot (1+\eta), & \text{if } U_d > \phi \\ B_d \cdot (1-\eta), & \text{if } U_d < (1-\phi) \\ B_d, & \text{otherwise} \end{cases} \tag{16}$$

where B_{d+1} is the budget for day $d+1$, U_d represents budget utilization rate on day d, $\phi = 0.90$ is the utilization threshold parameter, and $\eta = 0.10$ is the adjustment rate. This adaptive mechanism increases budget by 10% following high utilization days (> 90%), decreases budget by 10% following low utilization days (< 10%), and maintains stable allocation for moderate utilization levels. This enables budget self-calibration based on observed household response patterns over time.

4. Results

The initial daily budget B_0 of 10,000 flexpoints was determined through calibration via iterative simulation runs: the model was executed with varying B_0 values, and 10,000 was selected as the lowest value that consistently enabled full portfolio participation without systematic budget exhaustion. The exact starting value is less critical, as the adaptive budget mechanism (Eq. (16)) self-corrects based on realized utilization. Budget bounds were set at $0.5 \times B_0$ (lower) and $10 \times B_0$ (upper) to enable adaptive mechanism convergence while preventing excessive budget growth. Behavioral parameters follow household-specific thresholds ranging from 25 flexpoints (H7) to 145 flexpoints (H12), with participation sensitivity calibrated to capture realistic manual household response patterns. Table 2 summarizes the model input parameters.

The simulation was executed for 364 consecutive days spanning January through December 2023, using real 15-minute resolution smart meter data and historical day-ahead electricity prices from the ENTSO-E Transparency Platform [34]. Each daily optimization cycle processes up to 12 households sequentially through bilevel optimization problems solved using Gurobi 11.0.0 with a MILP formulation. Individual household optimizations converged within 1 to 5 s on average, with total daily computation time ranging from 15 to 45 s depending on household participation patterns. All computations were executed on a workstation equipped with an AMD Ryzen Threadripper PRO 5975WX processor (32 cores, 64 threads) and 512 GB RAM.

The sequential allocation mechanism processes households in hierarchical order, with HEMS households (H1–H6) prioritized before manual households (H7–H12) to ensure deterministic participation is allocated first. Within each category, household processing order is randomized daily as described in Section 3.3.5 to prevent systematic bias in budget access. Fig. 4 illustrates the sequential budget depletion pattern across the first two simulation days.

Day 1 begins with the initial budget of 10,000 flexpoints. HEMS households demonstrate consistent participation, with H5 consuming the largest allocation (1681 flexpoints) and H1 s (1129 flexpoints). Manual households exhibit heterogeneous participation patterns, with several households recording zero flexpoints due to non-participation. The day concludes with 2587 flexpoints remaining, representing 25.9% budget underutilization. Day 2 reflects the first adaptive budget adjustment, decreasing the starting allocation to 9000 flexpoints (10% reduction) in response to the previous day's low utilization rate. Participation patterns shift substantially: H1 earns 1017 flexpoints while H7 participates for the first time with an identical allocation. Despite this adjustment, 4480 flexpoints remain unutilized (49.8%), indicating that manual household participation uncertainty creates persistent budget inefficiency that the adaptive mechanism attempts to correct through iterative calibration.

Fig. 5 extends the analysis through Days 3–5, demonstrating the bidirectional adaptive mechanism response to varying utilization rates. Budget adjustment continues downward through Day 3 (8100 flexpoints) and Day 4 (7290 flexpoints) as the mechanism responds to sustained underutilization. Day 4 represents a critical transition point: despite the reduced budget, household participation converges to achieve near-complete utilization with only 1 flexpoint remaining. This triggers the first upward adjustment, increasing Day 5 budget to 8019 flexpoints (10% increase), validating the bidirectional nature of the adaptive mechanism. The observed pattern demonstrates that the framework successfully calibrates budget allocation in response to realized participation dynamics. This adaptive behavior persists throughout the full 364-day simulation period, with daily budgets continuously adjusting based on household response patterns.

While budget allocation dynamics demonstrate the framework's adaptive coordination capabilities, the fundamental objective of the DR program is to shift consumption from high-price to low-price periods. Fig. 6 illustrates aggregate household consumption patterns across the same five-day period, overlaying baseline consumption, optimized consumption, and electricity prices.

The aggregated load profile demonstrates consistent price-responsive behavior throughout the observation period. During price peaks, household consumption systematically falls below baseline levels as flexible loads are deferred. Conversely, during low-price periods and negative pricing events, household consumption increases substantially above baseline as deferred flexible loads execute during economically favorable windows. This validates that the bilevel optimization framework successfully translates reward signals into tangible load shifting behavior, achieving the core DR program objective of aligning consumption patterns with price signals.

Having established the operational mechanics through short-term dynamics, the annual performance metrics reveal substantial heterogeneity in household-level outcomes across the full 364-day simulation. Fig. 7 presents three key metrics aggregated over the complete simulation period: total cost savings, total flexpoints earned, and total energy shifted.

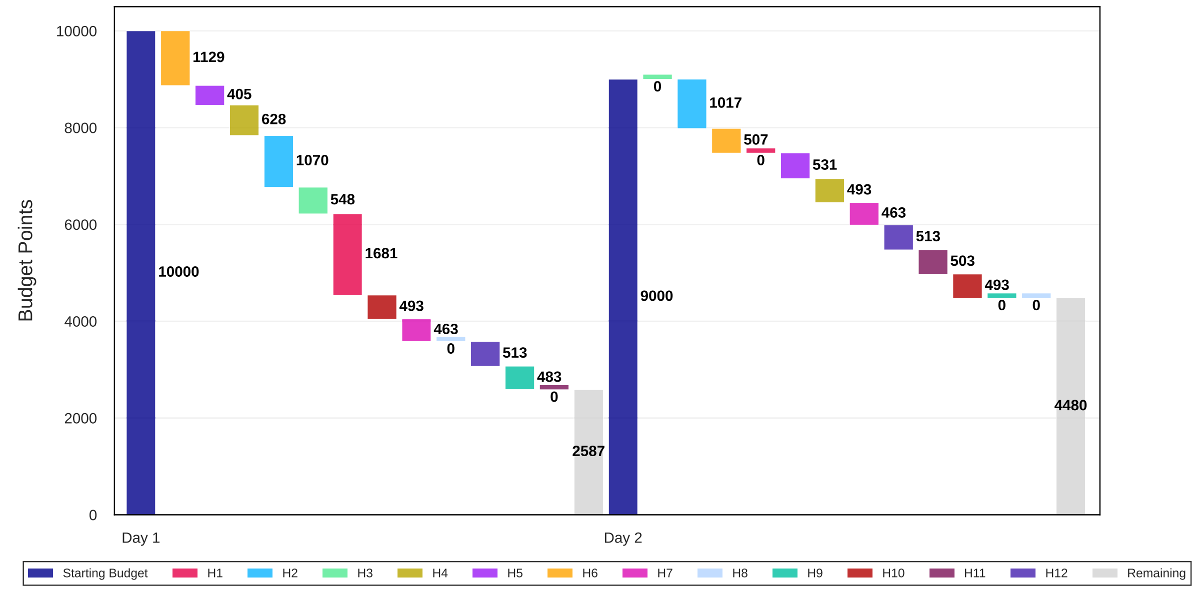

Fig. 4. Daily Budget Analysis: Days 1–2.

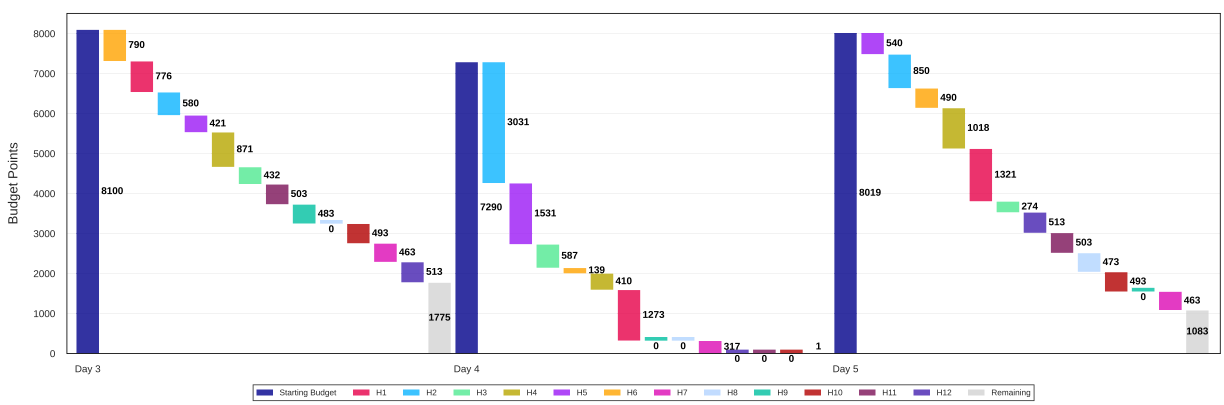

Fig. 5. Weekly Budget Allocation and Utilization Analysis.

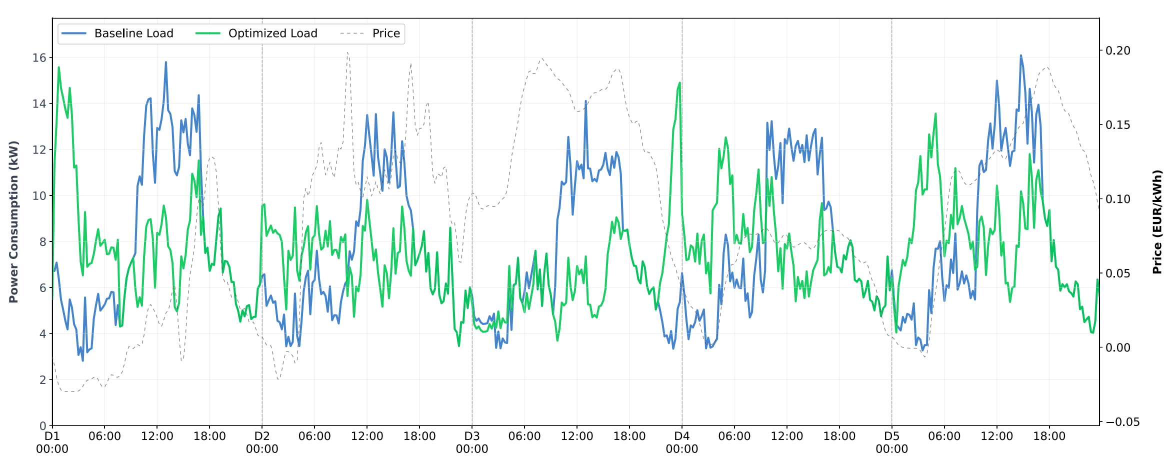

Fig. 6. 5-Day Load Profile Analysis.

Fig. 7 reveals substantial performance heterogeneity across the household portfolio. HEMS households (H1–H6) generally outperform manual households (H7–H12), with the best HEMS performer (H5) generating approximately three times the cost savings of the best manual performer (H7). However, performance rankings differ dramatically across the three metrics. In cost savings, H5 leads (228.5 EUR), followed by H2, H4, and H1. For energy shifted, H1 dominates (4539.9 kWh), with H2, H4, and H5 following. Flexpoints earned

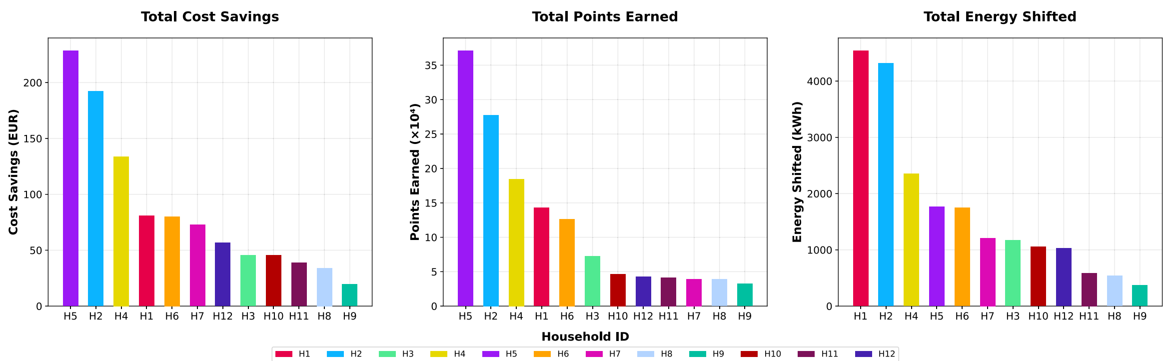

Fig. 7. Annual household performance metrics across 364 simulation days. Three critical insights emerge: (1) H1 shifts most energy yet ranks fourth in cost savings and flexpoints earnings, (2) H5 achieves highest cost savings and flexpoints while shifting less energy than H1, and (3) H7 (manual) outperforms H3 (HEMS) in cost savings despite fewer flexpoints.

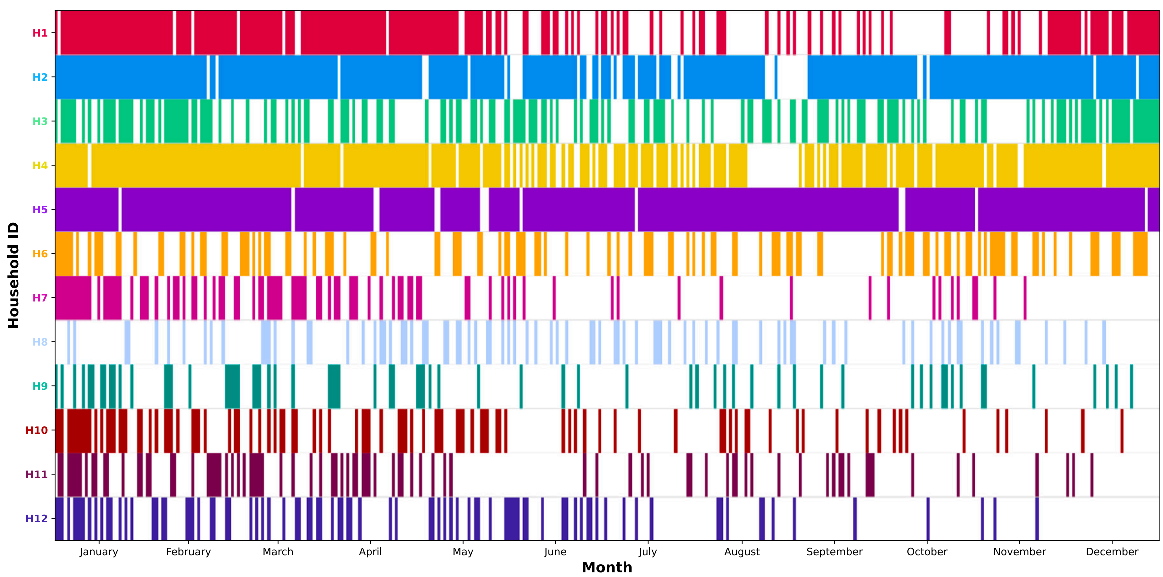

Fig. 8. Daily household participation patterns across 364 simulation days. Colored cells indicate days with flexpoint allocations; white cells represent non-participation. HEMS households (H1–H6) demonstrate dense, consistent participation throughout the year, with H2, H4, and H5 showing near-continuous engagement. Manual households (H7–H12) exhibit sparse, sporadic participation with visibly declining density as the year progresses, validating the willingness decay mechanism.

mirrors cost savings rankings closely. Among manual households, H7 substantially outperforms peers across all metrics, while H9 records the weakest performance. Notably, H7 achieves higher cost savings than H3 (HEMS) despite earning fewer flexpoints, demonstrating that performance depends on factors beyond simple participation frequency.

The underlying behavioral mechanisms driving these performance paradoxes become apparent when examining daily participation patterns. Fig. 8 presents a temporal heatmap showing participation status for all 12 households across 364 days, with colored cells indicating days when households received flexpoint allocations and white cells representing non-participation.

The participation matrix reveals participation patterns that directly explain the performance paradoxes from Fig. 7. H5 outperforms all households through near-continuous dense participation throughout the year, reflecting both consistent availability of flexible loads and reliable automation. This enables H5 to accumulate value across numerous participation days. In contrast, H1 demonstrates substantial sporadic participation with extended gaps, particularly during summer months. Despite H1 shifting the most total energy when flexibility is available and engaged, these participation gaps limit cumulative opportunity capture, explaining its fourth-place ranking in cost savings. The H7 versus H3 paradox emerges from an unexpected pattern: H3 (HEMS) exhibits surprisingly sparse participation with frequent gaps despite automation, while H7 (manual) maintains more consistent behavioral engagement. H7's higher participation frequency relative to other manual households, combined with H3's unexpectedly low participation rate, enables the manual household to outperform the automated one. This occurs despite H7 earning fewer total flexpoints. These patterns validate that consistent participation frequency, driven by both flexibility availability and engagement mechanisms, determines economic performance more than individual event magnitude.

Having examined individual household dynamics and participation patterns, the critical question for DR program viability is the aggregate flexibility potential extracted from the portfolio. Table 3 presents the system-level performance across the complete 364-day simulation, quantifying the total economic value and energy flexibility mobilized through the bilevel optimization framework.

The 12-household portfolio mobilizes 20,694.7 kWh of annual energy flexibility, generating 1028.5 EUR in cost savings through strategic load shifting. These savings represent the difference between optimized and baseline scheduling, quantifying the value added by bilevel coordination relative to uncoordinated consumption. This demonstrates substantial flexibility potential exists within residential portfolios when

Table 3
Annual system performance summary.

Household	Savings (EUR)	Flexpoints	Shifted (kWh)
H1	80.8	142,824	4539.9
H2	192.2	277,694	4320.0
H3	45.6	72,623	1170.5
H4	133.9	184,541	2354.0
H5	228.5	371,023	1768.9
H6	80.2	126,469	1750.4
HEMS Subtotal	761.1	1,175,174	15,903.6
H7	72.8	39,463	1210.5
H8	34.2	39,024	537.3
H9	19.5	32,345	371.5
H10	45.4	46,785	1055.4
H11	38.8	41,577	587.1
H12	56.7	43,093	1029.3
Manual Subtotal	267.4	242,287	4791.1
System Total	**1028.5**	**1,417,461**	**20,694.7**

appropriate coordination mechanisms are deployed. However, the contribution distribution reveals critical insights for portfolio composition strategy: HEMS households provide 74% of total cost savings (761.1 EUR) and 77% of energy shifted (15,903.6 kWh) despite representing only half the portfolio. This disproportionate contribution from automated households validates that automation substantially amplifies flexibility extraction beyond simple participation encouragement. For aggregators designing residential DR programs, these results suggest that portfolio composition significantly impacts economic viability, with technology adoption rates directly determining achievable flexibility volumes and program revenue potential. Monthly disaggregated results are provided in Appendix B.

5. Discussion

The value of flexibility and participation economics - The participation frequency insight reveals important considerations for DR program design and household compensation. For aggregators designing incentive mechanisms, rewarding consistent participation frequency yields greater portfolio value than incentivizing large individual load shifts.

This has direct implications for engagement mechanism design, where small but frequent rewards for regular participation outperform larger bonuses tied to energy volume shifted. Frequency-based compensation structures include: base payments for participation frequency (regardless of magnitude), tiered reward systems that increase with consecutive participation days, or reliability bonuses that reward households maintaining consistent engagement across weekly or monthly periods. The effectiveness of these mechanisms depends critically on the valuations of flexpoints, which aggregators determine based on revenue captured from flexibility market participation. Aggregators must balance flexpoint allocation to incentivize participation with the demonstrated portfolio performance providing concrete benchmarks for viable compensation levels. The adaptive budget mechanism parameters (±10% adjustment rate, 90% utilization threshold) were assumed here, but operational deployment requires calibrating these values to actual flexpoint costs and market revenue opportunities.

Flexible household portfolio composition strategy - The disproportionate contribution from HEMS households validates that automation substantially increases flexibility provision, compared to manual households. However, performance varies dramatically even within HEMS: H5 and H3 both are automated, yet H5 sees five times the cost savings. Aggregators recruiting households must assess characteristics beyond automation status, though identifying predictive screening criteria remains an open question for future research. The 50/50 HEMS/manual split demonstrated here represents a methodological assumption. Real portfolios vary widely in automation rates depending on market conditions and policy incentives. Aggregators should calibrate the HEMS/manual ratio to match actual automation penetration in their target group of households before applying the framework. Economic viability favors HEMS-heavy portfolios, as automation's reliable participation justifies higher recruitment costs given the 74% value contribution demonstrated. Modeling dynamic portfolio evolution where high-performing manual households graduate to HEMS status would capture realistic adoption pathways. DR program success often motivates households to invest in automation, creating feedback loops between engagement and technology adoption that static models miss. The framework thus enables aggregators to identify households with the highest flexibility potential from smart meter data alone, providing an evidence base for targeted HEMS deployment before any hardware investment.

Scalability of the proposed framework - The 12-household portfolio was selected to enable granular analysis of individual participation patterns while maintaining consumption diversity across the sample. This scale permitted identification of household-level dynamics that would be obscured in larger aggregations. Since each household is processed through an independent bilevel optimization, the framework is directly applicable to operational portfolios of thousands of households, enabling multi-year historical simulations for validation and behavioral parameter calibration.

Advantages and disadvantages of the proposed method - The framework operates on widely available 15-minute smart meter data without requiring submetering, appliance-level monitoring, or household preference data. It captures behavioral heterogeneity between automated and manual households, validated on real consumption data rather than synthetic profiles, and computation scales linearly with household count. However, the method is not suited for all contexts. When appliance-level control is required, submetering or NILM-based approaches are more appropriate, as the framework does not identify which specific devices are flexible. The baseline estimation requires a minimum of three months of smart meter history, making it inapplicable to new customers or recently relocated households. When higher-resolution data is available, disaggregation methods will yield more precise flexibility estimates than the aggregate approach proposed here.

Limitations and future framework extensions - The German household dataset provides smart meter data, but applying the framework across diverse electricity markets, climate zones, and building typologies would establish more generalizable findings. The minimum threshold filter removes small flexible loads to ensure practical DR activation, but discards potentially aggregatable flexibility. Additionally, the single continuous flexible period assumption excludes households that could provide multiple shorter flexibility events throughout the day. Behavioral parameters (participation thresholds, sensitivity, decay rates) and decomposition parameters (confidence adjustment factor, minimum flexibility threshold) represent assumed values requiring empirical calibration for operational deployment in specific market contexts. Field pilots with operational DR programs where real user engagement data can be collected would enable validation and refinement of these parameters. More comprehensive modeling of willingness decay mechanisms beyond the sigmoid-based participation probability implemented here would improve behavioral realism. Future validation efforts should compare the estimated flexibility against submetered data or NILM-based load disaggregation to assess whether identified flexible chunks align with truly shiftable loads. Field deployment with real DR activation would further enable comparison between estimated and realized flexibility.

6. Conclusion

Existing bilevel frameworks for residential demand response assume known flexibility characteristics, limiting their applicability to real-world portfolios where household flexibility must be estimated from available data. This work developed an appliance-agnostic, data-driven bilevel optimization framework that estimates flexibility directly from smart meter data and determines optimal reward allocations while accounting for behavioral heterogeneity across households.

For this portfolio, participation frequency determined economic performance more strongly than individual event magnitude, though further replications across diverse contexts are needed to confirm generalizability. Households with consistent participation accumulate greater value than those shifting larger energy volumes sporadically. HEMS households contributed the majority of total cost savings despite comprising half the portfolio, demonstrating that automation substantially amplifies flexibility extraction. These results support the economic case for HEMS adoption, as the demonstrated savings recur annually and the framework enables aggregators to quantify expected returns before any hardware investment. Within-category performance variance shows that technology type alone does not guarantee outcomes, requiring aggregators to assess household characteristics beyond automation status when recruiting portfolios.

For practitioners designing operational DR programs, these findings indicate that program value depends not only on the amount of flexibility households provide, but also on sustained participation over time, underscoring user retention as a critical design consideration. Future work should validate the flexibility extraction methodology against NILM-based approaches using field data from operational DR programs, and extend the framework to model dynamic portfolio evolution where manual households transition to automated participation over time.

CRediT authorship contribution statement

Reda El Makroum: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sebastian Zwickl-Bernhard:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **Lukas Kranzl:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **Hans Auer:** Writing – review & editing, Writing – original draft, Supervision.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude in order to refine writing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Reda El Makroum reports financial support was provided by Horizon Europe. Reda El Makroum reports financial support was provided by Swiss State Secretariat for Education Research and Innovation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the European Climate, Infrastructure and Environment Executive Agency (CINEA) through the Horizon Europe project PEDvolution under Grant Agreement No. 101138472. The authors acknowledge TU Wien Bibliothek for financial support for publishing through its Open Access Funding Programme.

Appendix A. Flexibility decomposition results

Table A.4 presents the numerical results of the appliance-agnostic flexibility decomposition for all 12 households. Total consumption is decomposed into non-flexible baseline and two flexible components: the single largest continuous flexible period per day retained for aggregation (Flex. cont.), and the remaining fragmented flexible load discarded by the single-period filter (Flex. frag.). Flexibility ranges from 16.2% (H1) to 47.4% (H4), with a portfolio average of 30.4%.

Table A.4
Annual flexibility decomposition per household.

HH	Total (kWh)	Non-flex. (kWh)	Flex. (cont.) (kWh)	Flex. (frag.) (kWh)	Total flex. (kWh)	Flex. (%)
H1	5602	4692	546	364	909	16.2
H2	7860	5636	1334	889	2223	28.3
H3	2143	1502	385	256	641	29.9
H4	2585	1359	736	490	1226	47.4
H5	7008	4405	1562	1041	2603	37.1
H6	2852	1809	626	417	1044	36.6
H7	7031	4757	1364	909	2274	32.3
H8	2482	1737	447	298	745	30.0
H9	2096	1633	278	185	463	22.1
H10	3713	2312	841	561	1402	37.7
H11	2827	2018	486	324	810	28.6
H12	6500	4845	993	662	1655	25.5
Total	**52,699**	**36,705**	**9597**	**6397**	**15,995**	**30.4**

Appendix B. Numerical simulation results

Table B.5 shows the results of the simulation for the year 2023 for each month.

Table B.5
Monthly aggregated simulation results (2023).

Month	Savings (EUR)	kWh shifted	Flexpoints
January	140.9	2907.5	172,912
February	103.8	2159.1	137,867
March	112.2	2314.2	126,456
April	92.8	2005.4	129,945
May	85.5	1869.5	101,474
June	67.4	1287.3	97,160
July	64.7	1202.2	97,745
August	67.3	1224.9	104,610
September	64.0	1164.1	79,548
October	77.1	1379.5	102,988
November	73.0	1418.9	121,888
December	79.7	1762.2	144,868
Total	**1028.5**	**20,694.7**	**1,417,461**

Data availability

The data that has been used is confidential.

References

[1] Golmohamadi H, Golestan S, Sinha R, Bak-Jensen B. Demand-side flexibility in power systems, structure, opportunities, and objectives: A review for residential sector. Energies 2024;17(18):4670. http://dx.doi.org/10.3390/en17184670, URL https://www.mdpi.com/1996-1073/17/18/4670.

[2] Yang S, Gao HO, You F. Demand flexibility and cost-saving potentials via smart building energy management: Opportunities in residential space heating across the US. Adv Appl Energy 2024;14:100171. http://dx.doi.org/10.1016/j.adapen.2024.100171.

[3] Li J, Ho MS, Xie C, Stern N. China's flexibility challenge in achieving carbon neutrality by 2060. Renew Sustain Energy Rev 2022;158:112112. http://dx.doi.org/10.1016/j.rser.2022.112112.

[4] Mahmood M, Chaudry S, Hussain A. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. Energy Convers Manag: X 2024;24:100790. http://dx.doi.org/10.1016/j.ecmx.2024.100790.

[5] Liu D, Qin Z, Hua H, Ding Y, Cao J. Incremental incentive mechanism design for diversified consumers in demand response. Appl Energy 2023;329:120240. http://dx.doi.org/10.1016/j.apenergy.2022.120240.

[6] Schirmer PA, Mporas I. Non-intrusive load monitoring: A review. IEEE Trans Smart Grid 2023;14(1):769–84. http://dx.doi.org/10.1109/TSG.2022.3189598.

[7] Brudermueller T, Breer F, Staake T. Disaggregation of heat pump load profiles from low-resolution smart meter data. In: Proceedings of the 10th ACM international conference on systems for energy-efficient buildings, cities, and transportation. Association for Computing Machinery; 2023, p. 228–31. http://dx.doi.org/10.1145/3600100.3623731.

[8] Huy THB, Dinh HT, Kim D. Real-time energy scheduling for home energy management systems with an energy storage system and electric vehicle based on a supervised-learning-based strategy. Energy Convers Manage 2023;292:117340. http://dx.doi.org/10.1016/j.enconman.2023.117340.

[9] Yin C, Dong J, Zhang Y. Distributionally robust bilevel optimization model for distribution network with demand response under uncertain renewables using Wasserstein metrics. IEEE Trans Sustain Energy 2025;16(2):1165–76. http://dx.doi.org/10.1109/TSTE.2024.3509314, URL https://ieeexplore.ieee.org/document/10771664.

[10] Parrish B, Heptonstall P, Gross R, Sovacool BK. A systematic review of motivations, enablers and barriers for consumer engagement with residential demand response. Energy Policy 2020;138:111221. http://dx.doi.org/10.1016/j.enpol.2019.111221.

[11] Miri M, McPherson M. Demand response programs: Comparing price signals and direct load control. Energy 2024;288:129673. http://dx.doi.org/10.1016/j.energy.2023.129673, URL https://www.sciencedirect.com/science/article/pii/S0360544223030670.

[12] Thrän J, Green TC, Shorten R. Levelised cost of demand response: Estimating the cost-competitiveness of flexible demand. Energy Convers Manage 2025;349:120811. http://dx.doi.org/10.1016/j.enconman.2025.120811.

[13] Papias I, Michalakopoulos V, Sarmas E, Marinakis V. Aggregated flexibility dynamic forecasting of residential energy prosumers for DR programs. In: 2024 15th international conference on information, intelligence, systems & applications. 2024, p. 1–9. http://dx.doi.org/10.1109/IISA62523.2024.10786677, URL https://ieeexplore.ieee.org/document/10786677.

[14] Papias I, Michalakopoulos V, Sarmas E, Marinakis V, Antonesi G, Cioara T, Anghel I. A data-driven framework for estimating residential energy flexibility for aggregated demand-side management. Sustain Energy, Grids Networks 2025;43:101783. http://dx.doi.org/10.1016/j.segan.2025.101783, URL https://www.sciencedirect.com/science/article/pii/S2352467725001651.

[15] Zhang W, Wu J, Liu J. A novel building flexibility potential assessment method based on hybrid CNN-GRU-CALDA framework considering consumer psychology. Sustain Cities Soc 2024;104:105298. http://dx.doi.org/10.1016/j.scs.2024.105298.

[16] Bampoulas A, Pallonetto F, Mangina E, Finn DP. A Bayesian deep-learning framework for assessing the energy flexibility of residential buildings with multicomponent energy systems. Appl Energy 2023;348:121576. http://dx.doi.org/10.1016/j.apenergy.2023.121576.

[17] Haapaniemi J, Haakana J, Räisänen O, Tikka V, Lassila J, Rautiainen A. Quantification of implicit price flexibility of household customers' load demand with machine learning and Shapley analysis. Energy 2025;332. http://dx.doi.org/10.1016/j.energy.2025.136980, URL https://ideas.repec.org/a/eee/energy/v332y2025ics0360544225026222.html.

[18] Plaum F, Ahmadiahangar R, Rosin A, Kilter J. Aggregated demand-side energy flexibility: A comprehensive review on characterization, forecasting and market prospects. Energy Rep 2022;8:9344–62. http://dx.doi.org/10.1016/j.egyr.2022.07.038.

[19] Petrucci A, Vallianos C, Candanedo JA, Buonomano A, Athienitis A. Large-scale load profiling for energy flexibility in residential buildings: A data-driven approach for load aggregation through representative clusters. Energy Build 2025;346:116195. http://dx.doi.org/10.1016/j.enbuild.2025.116195, URL https://www.sciencedirect.com/science/article/pii/S0378778825009259.

[20] Kallel S, Amayri M, Bouguila N. Clustering and interpretability of residential electricity demand profiles. Sensors 2025;25:2026. http://dx.doi.org/10.3390/s25072026.

[21] Vahedi S, Zhao L, Tanjim N, Lee W-J. Data-driven classification for residential coincident peak demand contributors using actual power, sociological, and meteorological data. IEEE Trans Ind Appl 2024;60:2443–52. http://dx.doi.org/10.1109/TIA.2023.3343673.

[22] Wang H, Guo F, Sivakumar A. Analysing the impact of electric vehicle charging on households: An interrelated load profile generation approach. Energy Build 2025;335:115558. http://dx.doi.org/10.1016/j.enbuild.2025.115558, URL https://www.sciencedirect.com/science/article/pii/S0378778825002889.

[23] Zhang W, Wu J. Optimal real-time flexibility scheduling for community integrated energy system considering consumer psychology: A cloud-edge collaboration based framework. Energy 2025;320:135340. http://dx.doi.org/10.1016/j.energy.2025.135340, URL https://www.sciencedirect.com/science/article/pii/S036054422500982X.

[24] Sinha A, Malo P, Deb K. A review on bilevel optimization: From classical to evolutionary approaches and applications. IEEE Trans Evol Comput 2018;22:276–95. http://dx.doi.org/10.1109/TEVC.2017.2712906.

[25] Maharjan S, Zhu Q, Zhang Y, Gjessing S, Basar T. Dependable demand response management in the smart grid: A Stackelberg game approach. IEEE Trans Smart Grid 2013;4:120–32. http://dx.doi.org/10.1109/TSG.2012.2223766.

[26] Zhang J, Li S, Li L, Zu G, Wang Y, Yang T. A bi-level method for flexibility feature extraction and user clustering based on real-world data from independent smart meters of residential electric vehicle users. Energies 2025;18(4):873. http://dx.doi.org/10.3390/en18040873, Publisher: Multidisciplinary Digital Publishing Institute, URL https://www.mdpi.com/1996-1073/18/4/873.

[27] Wang Y, Yan Y, Lin Z, Zhang Z, Ma Y, Yang L, Chen X, Yu K. Score-based incentive demand response for load aggregators considering power-score redeem behavior of residential consumers. Int J Electr Power Energy Syst 2024;162:110233. http://dx.doi.org/10.1016/j.ijepes.2024.110233, URL https://www.sciencedirect.com/science/article/pii/S014206152400454X.

[28] Lin J, Dong J, Liu D, Zhang Y, Ma T. From peak shedding to low-carbon transitions: Customer psychological factors in demand response. Energy 2022;238:121667. http://dx.doi.org/10.1016/j.energy.2021.121667.

[29] Barala CP, Firdous A, Mathuria P, Bhakar R. Bi-level framework for coordination between thermostatically controlled loads and distribution system for flexibility. Electr Power Syst Res 2024;230:110221. http://dx.doi.org/10.1016/j.epsr.2024.110221, URL https://www.sciencedirect.com/science/article/pii/S0378779624001093.

[30] Dong Y, Wuken E, Zhang H, Ren P, Zhou X. Bi-level coordinated operation optimization of multi-park integrated energy systems considering categorized demand response and uncertainty: A unified adaptive robust optimization approach. Renew Energy 2025;241:122331. http://dx.doi.org/10.1016/j.renene.2024.122331, URL https://www.sciencedirect.com/science/article/pii/S0960148124023991.

[31] Davoudi M, Moeini-Aghtaie M, Mehrtash M. Optimal operation of a residential energy hub participating in electricity and heat markets. Renew Energy Focus 2024;51:100646. http://dx.doi.org/10.1016/j.ref.2024.100646, URL https://www.sciencedirect.com/science/article/pii/S1755008424001108.

[32] Rasouli V, Gomes Á, Antunes CH. A genetic algorithm approach for aggregation of residential electricity prosumers' flexibility. In: García-Sánchez P, Hart E, Thomson SL, editors. Applications of evolutionary computation. Cham: Springer Nature Switzerland; 2025, p. 435–51. http://dx.doi.org/10.1007/978-3-031-90062-4_27.

[33] Shi R, Jiao Z. Individual household demand response potential evaluation and identification based on machine learning algorithms. Energy 2023;266:126505. http://dx.doi.org/10.1016/j.energy.2022.126505, URL https://www.sciencedirect.com/science/article/pii/S0360544222033916.

[34] Hirth L, Mühlenpfordt J, Bulkeley M. The ENTSO-E transparency platform – A review of Europe's most ambitious electricity data platform. Appl Energy 2018;225:1054–67. http://dx.doi.org/10.1016/j.apenergy.2018.04.048.

[35] Kazmi H, Mehmood F, Shah M. Quantifying residential energy flexibility potential for demand response programs using observational data from grid outages: Evidence from Pakistan. Energy Policy 2024;188:114094. http://dx.doi.org/10.1016/j.enpol.2024.114094.

[36] Azizi E, Ahmadiahangar R, Rosin A, Martins J, Lopes RA, Beheshti M, Bolouki S. Residential energy flexibility characterization using non-intrusive load monitoring. Sustain Cities Soc 2021;75:103321. http://dx.doi.org/10.1016/j.scs.2021.103321.

[37] Chen M, Fang K, Huang Q, Yang S, Cheng H, Liu T, Huang Y. A triple-threshold-based load event detection algorithm for non-intrusive load monitoring. In: Journal of physics: conference series, vol. 2065, 2021, 012011. http://dx.doi.org/10.1088/1742-6596/2065/1/012011.