DOI: https://doi.org/10.15276/aait.09.2026.36
UDC 004.62:621.31

A data mining method for non-intrusive load monitoring in proactive energy consumption systems for smart facilities

Andrii Ya. Zasidko[1)]
ORCID: https://orcid.org/0000-0001-5197-0435; andzasidko@op.edu.ua. Scopus Author ID: 57792636800
Olena O. Arsirii[1)]
ORCID: https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900
[1)] Odesa Polytechnic National University, 1 Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

A smart facility needs an estimate of its future peak load before it can allocate power for a planned operation. This study develops a data mining method for non-intrusive load monitoring in proactive energy consumption systems for smart facilities. The method combines data preprocessing, operational event extraction and predictive modeling in a single pipeline. Using the REFIT dataset, it converts raw aggregate power time series into discrete operational events and their attributes (onset times, durations and daily usage profiles), quantitative estimates of recognition accuracy and a forecast-based estimate of quantile-based power headroom for proactive control. We compare two-state combinatorial optimization with a causal engineering detector of power changes. The experiments demonstrate that rankings by power reconstruction error and event recognition can differ. In the final stage, we add the extracted appliance-level features to a regularized linear regression model, estimate error quantiles on the validation set and calculate quantile-based power headroom. Detector features reduce the mean absolute error of five-minute peak forecasts by zero point four, one point four and four point eight watts in the three homes. The detector has higher event F_1 in all six house–appliance pairs. The resulting estimate gives a proactive controller a basis for checking how much power remains available over the forecast horizon.

Keywords: Data mining; non-intrusive load monitoring; time-series disaggregation; operational event modeling; time-series forecasting; proactive energy consumption

For citation: **Zasidko A. Ya., Arsirii O. O. "A data mining method for non-intrusive load monitoring in proactive energy consumption systems for smart facilities". *Applied Aspects of Information Technology.* 2026; Vol.9 No.4: 561–573. DOI: https://doi.org/10.15276/aait.09.2026.36**

INTRODUCTION

To allocate resources in advance and control energy use proactively in a smart facility, we need to forecast the load at its shared service entrance throughout the planned operation. This entrance is the facility's single connection to the external grid and is subject to a power limit. An instantaneous reading from the main meter cannot capture the future maximum over that interval. A high-power local load, such as a cycling heater or another energy-intensive appliance, may switch on unexpectedly. Conventional submetering requires a separate telemetry channel for each device, making it costly and difficult to deploy. Non-intrusive load monitoring (NILM) addresses this constraint by estimating individual appliance contributions from a single aggregate measurement [2], [7].

A low average power reconstruction error across all time samples gives little assurance that a NILM method identifies when high-power local loads start. Appliances remain off most of the time, so the samples are highly imbalanced. Errors in a few brief, high-amplitude activations, including missed activations, have little effect on the average. Baseline NILM methods such as two-state combinatorial optimization (CO) and simple edge detectors (EDGE) can generate false detections or miss local states, respectively. This difference separates power reconstruction accuracy from event recognition performance. In proactive energy consumption systems, missed or false starts distort event counts and their temporal profiles. NILM therefore needs data mining methods that extract operational events accurately, preserve their timing patterns and improve forecasts of available power.

LITERATURE REVIEW AND PROBLEM STATEMENT

Research on data mining for NILM follows three related lines, from classical stochastic disaggregation to the use of disaggregated loads in proactive energy forecasting.

Hart [7] established the basis for NILM by formulating aggregate load disaggregation in terms of appliance states. Later work introduced probabilistic models and evaluation tools. Batra et al. [2] proposed NILMTK, an open-source toolkit that standardized the evaluation of disaggregation algorithms. Lucas et al. [18] used combinatorial optimization (CO) and a factorial hidden Markov model (FHMM) in a framework for estimating residential load flexibility. CO selects the

© Zasidko A., Arsirii O., 2026

This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/deed.uk)

combination of appliance states that minimizes the difference between measured aggregate power and the sum of individual appliance power levels [2], [18]. An additive FHMM represents each appliance as a separate Markov chain with hidden states such as off, standby and active. The observed signal is the sum of their outputs [14]. In load flexibility applications, whether an appliance can be rescheduled depends on additional empirical assumptions about user comfort [18]. Those assumptions cannot be derived from the initial disaggregation alone, and errors can accumulate when short operational cycles are identified.

As data mining methods developed, research turned to machine learning algorithms capable of processing continuous time series. Kelly and Knottenbelt [9] were among the first to apply deep neural networks systematically to energy disaggregation. Subsequent work used sequence-to-point learning [26]. Deterministic approaches developed in parallel. Engineering edge detectors (EDGE), for example, detect abrupt changes in the aggregate signal. They use the first difference or rate of change of aggregate power ($\Delta P(t)$) to identify rising edges associated with activation and falling edges associated with deactivation [7], [22], [27]. Event-based methods then assign detected transitions to appliances using characteristics such as power levels and operating patterns [1]. A model can achieve low average reconstruction error (MAE or RMSE) while still missing brief, high-power appliance activations. Comparisons of NILM algorithms therefore report both power reconstruction and event or state recognition metrics [3], [23]. Welikala et al. [24] used state durations and time-of-day patterns to refine NILM results and forecast consumption five minutes ahead. Their analysis used aggregate measures from several households, whereas the present study evaluates performance at each smart facility. NILM-Flex [5] also links disaggregated loads to a behavioral model of flexibility. Detailed delays and the preservation of start frequencies lie outside its scope.

Li and Sun [16] evaluate a sequence-to-point network on REFIT, including the kettle and microwave. Their model uses regularly resampled data and a centered window, and their F_1 counts operating-state samples. Our F_1 counts one-to-one onset matches on the native irregular grid. These scores measure different outcomes and use different house splits. Li et al. [17] use disaggregated histories to forecast individual appliance loads and flexible demand on UK-DALE. The present forecast target is the five-minute maximum of the aggregate load in each house.

The studies reviewed here link NILM, usage patterns and forecasting. Much of the work focuses on average power reconstruction error or on detailed models of human behavior. The relationship between accurate extraction of discrete operational events (onset, duration and frequency) and the quality of forecasting features for proactive control remains unresolved. This motivates a data mining method that works without subjective behavioral assumptions.

RESEARCH AIM AND OBJECTIVES

The object of this study is non-intrusive monitoring of electrical loads in proactive energy consumption systems for smart facilities. Its subject is the data mining models and methods used in these systems.

The aim is to develop a method for mining NILM data that preserves the frequency, onset times and durations of operationally defined activations and improves forecasts of the future aggregate maximum using appliance-level features.

The study addresses the following objectives.

1) Obtain heterogeneous aggregate and appliance-level consumption records from an open repository and preprocess observations collected under varying calendar conditions.

2) Analyze local time-series dynamics to calculate statistical descriptors and incorporate calendar context.

3) Convert the continuous power signal into discrete states and transitions using EDGE or CO with event confirmation rules based on thresholds and time delays to suppress false detections.

4) Match detected onsets to reference labels one to one, assess how well daily onset profiles are recovered and compare rankings based on power MAE and event F_1.

5) Combine the extracted appliance-level features with aggregate history at multiple timescales in ridge regression, estimate validation error quantiles and calculate quantile-based power headroom for proactive energy consumption.

MATERIALS AND METHODS

Fig. 1 shows the processing pipeline. The first 27 calendar days are used to fit the disaggregation signatures and the next 27 to fit the forecast model. The following 18 days are used for validation and quantile calibration, and the final 18 for testing. Hysteresis and time confirmation follow disaggregation and precede feature construction and event evaluation. The model receives the appliance estimates as separate predictors alongside aggregate power.

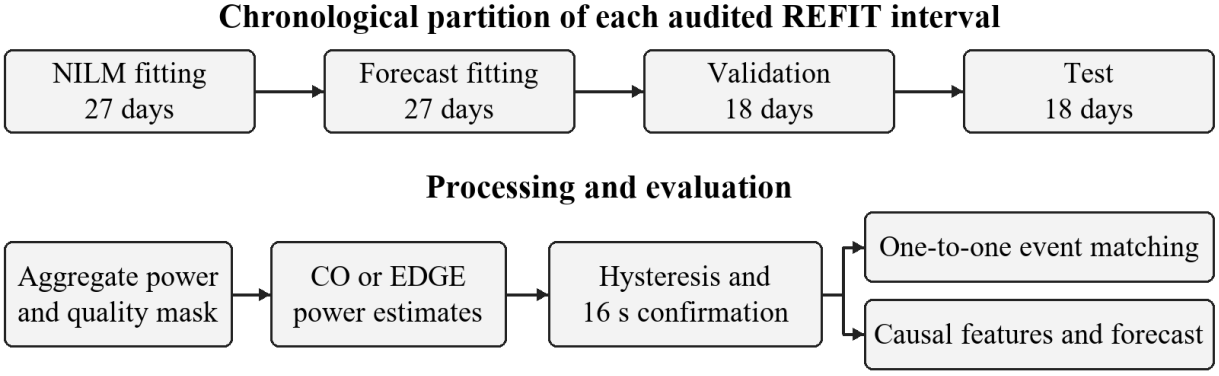


Fig. 1. **Processing stages from data audit to event evaluation and peak-load forecasting**
Source: **compiled by the authors**

At the measurement boundary, we express aggregate power as a regression model with a deterministic and a random component. The deterministic component includes the power contributions of the appliances and the unmeasured background load, all in watts

$$P_{agg}(t) = \sum_{j=1}^{J} P_j(t) + P_{other}(t) + \varepsilon(t). \quad (1)$$

Here, t is the timestamp, $P_j(t)$ is the power of appliance j, and $P_{other}(t)$ is background power.

The zero-mean random term $\varepsilon(t)$ accounts for measurement discrepancies and model simplifications.

The forecast target is the maximum aggregate power recorded after time t and within a horizon of length H.

$$M_t = \max_{0<\delta\leq H} P_{agg}(t+\delta). \quad (2)$$

Equation (2) defines the target using actual observations at offsets δ with $0 < \delta \leq H$. We use H = 300 s and verify coverage of the future interval. All model inputs are available at the forecast origin.

REFIT data and chronological preprocessing

The study uses REFIT, an open dataset of aggregate and appliance-level household measurements [19], [20]. Other NILM benchmarks provide both measurement levels using household recordings [10] or simulation [12]. The REFIT sensors do not measure mains voltage, and the appliance transmitters are not synchronized [19]. Equation (1) therefore models the recorded quantities without assuming exact metrological balance and retains the sign of the residual $\varepsilon(t)$. The selected homes had no reported solar generation. Grid imports in a facility with generation or storage would require a different power balance.

We obtained the unchanged CSV files from the associated Zenodo record 5063428. We checked file sizes and MD5 checksums against the record and saved SHA-256 hashes in a checksum file. We verified the target channels against the metadata of the REFIT converter in NILMTK and the original dataset description. Houses 2, 4 and 5 contained the target appliance categories and had no reported solar generation. We excluded House 1, downloaded during the initial audit, because it lacked the target appliance information. Before running NILM, we fixed a 90-calendar-day interval for each selected house. Each interval had to contain at least 80 eligible days overall and at least 70 % eligible days in each chronological subset. The initial interval for House 2 failed this check and was replaced following the data quality audit. The first intervals for Houses 4 and 5 met the requirements.

The electric kettle and microwave oven were selected for their distinct load signatures. The kettle is a high-power resistive device with brief, steep rising edges. The microwave produces a pulsed profile as its magnetron cycles. These two load types allow the confirmation rules to be tested on both short single pulses and compound duty cycles.

Table 1. **Sample and eligible days by chronological subset**

House	Interval	Channels Kettle / microwave	Eligible days	Training / validation / test
2	[2014-03-06, 2014-06-04)	8 / 5	80/90	46 / 16 / 18
4	[2013-10-11, 2014-01-09)	9 / 8	84/90	50 / 18 / 16
5	[2013-09-26, 2013-12-25)	8 / 7	85/90	49 / 18 / 18

Source: **compiled by the authors**

Table 1 gives the half-open observation intervals $[t_1, t_2)$, the target appliance channels, the number of eligible days and their distribution across training, validation and test subsets. Data and intervals were selected independently of the disaggregation results.

Channel numbers follow the Appliance columns in the original CSV files. The chronological training, validation and test periods span 54, 18 and 18 calendar days, respectively.

We converted the original Unix timestamps in UTC to Europe/London local time, accounting for daylight saving time when deriving calendar features and daily boundaries. We retained the irregular sample times without averaging to one-minute intervals. A calendar day is eligible when coverage reaches at least 95% and measurements are available within two minutes of each local day boundary. When calculating coverage, at most 24 seconds are counted between adjacent timestamps. At the sample level, we exclude records with a nonzero Issues flag and the first sample after any gap longer than 120 seconds. An Issues code alone does not establish that a physical fault occurred. Reporting sampling and quality rules helps interpret comparisons across datasets with different measurement and annotation practices [11].

The REFIT cleaning specification limits how these records can be interpreted. Short gaps were filled with the previous value and long gaps with zeros. The CSV files do not reveal the provenance of every zero reading, so we use the cleaned appliance channel as an independent operational reference without treating every zero as conclusive evidence that the appliance was off. The methods are compared on a shared observation mask, with coverage reported separately.

We split the observation window chronologically into 54 training days, 18 validation days and 18 test days. Table 1 report the eligible days in each house. The first 27 days are used to learn the NILM signatures. The next 27 days supply training examples for the forecasting model using out-of-sample estimates from the fixed disaggregation algorithms. The split keeps reference appliance states out of the forecasting inputs and prevents data leakage. All history windows and future forecast horizons lie wholly within their respective subsets.

Rolling statistics and calendar features

To forecast the future load maximum M_t in (2) from the original measurements and baseline disaggregation estimates, we analyze the local dynamics of the time series. At each step t, we calculate the following statistics for aggregate power $P_{agg}(t)$ in (1) over a rolling window of size W.

The local moving mean estimates the trend.

$$\mu_t = \frac{1}{W}\sum_{i=0}^{W-1} P_{agg}\,(t-i), \qquad (3)$$

The local standard deviation measures volatility.

$$\sigma_t = \sqrt{\frac{1}{W}\sum_{i=0}^{W-1}\left(P_{agg}(t-i) - \mu_t\right)^2}, \qquad (4)$$

The local range A_t measures the amplitude of fluctuations.

$$A_t = \max_{0\le i<W} P_{agg}(t-i) - \min_{0\le i<W} P_{agg}(t-i). \qquad (5)$$

The model encodes local hour cyclically using sine and cosine terms.

$$h_{\sin}(t) = \sin\left(\frac{2\pi\cdot \mathrm{hour}(t)}{24}\right), h_{\cos}(t) = \cos\left(\frac{2\pi\cdot \mathrm{hour}(t)}{24}\right), \qquad (6)$$

Here, $\mathrm{hour}(t)$ is the local hour of day in Europe/London.

The forecast combines these rolling descriptors and calendar indicators with the confirmed appliance features when the corresponding variant is used.

Baseline disaggregation and event confirmation

The first baseline, CO, implements standard combinatorial optimization [2], [18]. We use deterministic one-dimensional clustering on the first training subset to estimate two power levels for each of the nine appliance channels [2]. The background term b corresponds to the dynamic unmodeled power $P_{other}(t)$ in (1). For optimization, it is held constant and set to the median training residual between aggregate power and the sum of the nine appliance channels. For each new sample, we select the combination of levels as follows.

$$\hat{c}(t) = \underset{c\in C}{arg\,min}\left|P_{agg}(t) - b - \sum_{j=1}^{9} c_j\right|. \qquad (7)$$

The set C contains up to 512 combinations of nine two-state models (K = 2). The selected level of channel j is c_j, and $P_{agg}(t)$ is the aggregate active power of the entire facility (house or apartment) at time t.

A fixed ordering breaks ties. Training is local and uses labeled appliance measurements. Deployment in a new house would require prior labeling. The decision for each sample is independent of future readings and previous states, which can lead to rapid switching between combinations. Transfer across domains has been

investigated separately for neural disaggregation methods [4].

We checked the code against the open NILMTK implementation [2], [3]. The standard package was not run as a black box because its two-state configuration differs substantially from our clustering settings. Fixing the implementation and using a shared protocol allows the comparison to be reproduced [3], [13], [23]. Unknown appliances and equipment with multiple operating modes limit the use of CO and make additional filtering rules necessary.

The second baseline, EDGE, is an engineering detector of changes in aggregate power. CO independently selects a fixed combination of states at each sample, while EDGE looks for changes in the aggregate signal. To identify possible switch-on and switch-off times at each step t, it calculates the difference in power over a specified interval.

$$\Delta P(t) = P_{agg}(t) - P_{agg}(t - \ell). \quad (8)$$

Here ℓ is a lag of one observed sample within a continuous segment.

The training appliance channels provide typical amplitudes for the target loads. A positive or negative change $\Delta P(t)$ is matched to a switch-on or switch-off event using a fixed relative tolerance of 25 %. If several matches are possible, the detector selects the one with the smallest relative discrepancy, using a fixed appliance order. The maximum time in the active state is set to the 0.99 quantile of complete event durations in the training data. At the start of each segment, the model assumes that the appliance is off and can therefore miss a load that is already active. EDGE is a simplified engineering detector based on the approach introduced by Hart [7]. Simultaneous transitions, similar signatures and missed switch-offs are expected limitations.

The same confirmation rule is applied to the reference and estimated appliance power series. An ON candidate begins at or above the upper level threshold and an OFF candidate at or below the lower one. These thresholds are 1000 and 500 W for the kettle and 200 and 100 W for the microwave. The condition must persist for at least 16 s in observed time. These levels apply to appliance power, while the 25 % relative tolerance applies to changes in aggregate power used by EDGE [21].

The level thresholds, persistence times, and relative tolerance were fixed engineering settings. The EDGE signatures and the duration cap were derived separately for each appliance from complete training events. The cap was the training duration quantile at level 0.99. The protocol did not include an optimization of the fixed settings.

These parameters are fixed before testing. The event onset is the first sample in the confirmed sequence. Detection occurs when confirmation is complete, and only then does the event become available to the forecast model.

Gaps longer than 24 seconds or invalid samples split the observations into separate segments. Activations that start before a segment or continue beyond its end are treated as censored and counted separately from complete events. For an active appliance, we track elapsed time from a known onset. An onset from an earlier segment does not carry over across a gap. Availability is determined by the algorithm's confirmation time. Actual network delays were outside the measurements.

Event matching and performance measures

We evaluate disaggregation by matching detected onsets one to one with reference labels using bipartite matching. The allowed time difference is 60 s, with a separate evaluation at 120 s. The procedure first maximizes the number of matched pairs and then minimizes the total timing error across those pairs. Each reference event can be counted only once.

For n reference events, m estimated events and k matched pairs, precision, recall and F_1 are defined by (9).

$$Precision = \frac{k}{m}, Recall = \frac{k}{n}, F_1 = \frac{2k}{n+m}. \quad (9)$$

A metric is undefined when its denominator is zero. Onset and duration errors are calculated only for matched pairs and are reported alongside the numbers of matches and misses. Power reconstruction is evaluated separately using mean absolute error (MAE) in watts.

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|P_j(t) - \hat{P}_j(t)\right|, \quad (10)$$

Here, $P_j(t)$ and $\hat{P}_j(t)$ are the reference and estimated power of appliance j at sample t, and N is the number of test samples. MAE summarizes power reconstruction across all samples. Short events require a separate detection metric.

We group onsets into 24 local hourly bins.

If N_{jb} is the number of onsets of appliance j in bin b, the normalized profile and the distance between profiles are given by (11).

$$p_{jb} = \frac{N_{jb}}{\sum_{b=0}^{23} N_{jb}}, \; D_{TV} = ½ \sum_{b=0}^{23} |\hat{p}_{jb} - p_{jb}|. \quad (11)$$

The profile describes when recorded starts occurred. It gives neither the probability of

occupancy nor the probability of at least one start in a given hour. The normalized profile is undefined when there are no events. We assess total onset counts and the mean absolute difference in daily counts separately, since identical normalized profiles can result from different usage frequencies.

Both methods are evaluated on the same valid test samples. In the post hoc analysis, we vary the reference segment gap limit across 24, 60 and 120 s on the same quality mask. We retain the original thresholds and compare complete cycles and their durations by onset timestamp. We form exploratory 95% percentile intervals for complete-cycle F_1 by resampling eligible local days, with 20,000 draws. The saved matched pairs have reference and estimated onsets on the same local day, so each day contributes its original reference, detection and match counts.

Peak-load forecasting and power headroom

Each forecast variant uses a separate ridge model for each house [8], [15], with the regularization coefficient fixed at 10. The baseline F_{base} uses aggregate and calendar predictors. F_{CO} and F_{EDGE} add causal appliance estimates from CO and EDGE. F_{oracle} uses the same features derived from recorded appliance channels as an information reference. The cleaned channels and the fixed linear model both limit its interpretation.

Forecast times are scheduled every 60 s and aligned to the latest valid sample. A retained origin requires continuous observations over the preceding hour and following five minutes, with no gap longer than 24 s. F_{base} uses 15 predictors: cyclic hour and weekday terms, a weekend indicator, current aggregate power, values lagged by 60, 300 and 900 s, and the mean and maximum over each of those durations. Each appliance adds a confirmed state, the number of confirmed starts in the previous hour, elapsed activation time, and the latest available power. Standardization is fitted on the forecast training subset.

For each variant, the upper estimate $\hat{U}_t$ is the point forecast plus the empirical 90th percentile of signed validation residuals (observed minus forecast peak). The nominal level of 0.90 illustrates a margin targeting one upper miss in ten forecasts under similar conditions. Given a fixed illustrative limit P_lim, we define quantile-based power headroom by (12).

$$R_t = P_{lim} - \hat{U}_t. \quad (12)$$

For the 5 kW illustration, a controller flags possible exceedance when estimated headroom is zero or negative. The actual service rating is facility-specific. For retrospective screening at 3, 4 and 5 kW, we count forecast origins at which the measured future maximum exceeds the limit. We report recall among those origins and false-positive rate among origins without an exceedance. Adjacent origins can cover the same physical peak.

For F_{EDGE} versus F_{base}, we resample paired absolute-error sums and origin counts by local day with 20,000 draws. The resulting percentile intervals are conditional on these houses and preserve dependence within days. We also tabulate MAE on origins with a recorded target-appliance onset inside the horizon. The subgroup is defined retrospectively using known reference onsets.

RESEARCH RESULTS

Event recognition and power reconstruction

Both methods returned estimates at every shared valid test sample, giving 100 % coverage of that series. The test set covers 18, 16 and 18 eligible days in Houses 2, 4 and 5, respectively. The independent appliance channels yielded 60 complete reference cycles, 29 for the kettle and 31 for the microwave. We calculated separate metrics for each house and appliance pair.

Table 2 reports k matched pairs, n reference cycles, and m estimated cycles at the primary tolerance of 60 s. D_{tv} is the total variation distance between hourly onset profiles. Figure 2 shows a six-hour excerpt from House 4 on 22 December 2013, starting one hour before its first confirmed reference kettle onset.

Table 2 and Fig. 2 show how CO and EDGE differ in power reconstruction and event recognition.

Event recognition. At the primary tolerance of 60 s, EDGE has a higher F_1 for complete cycles in all six house and appliance pairs. Kettle F_1 ranges from 0.129 to 0.560 for EDGE and from 0.032 to 0.170 for CO. For the microwave, it ranges from 0.089 to 0.605 for EDGE and from 0.013 to 0.079 for CO.

For the microwave in House 5, CO has the lower power MAE (52.1 versus 84.6 W), while EDGE has the higher F_1 (0.089 versus 0.013). The kettle in the same house has a numerically reversed MAE ranking as well, but its MAEs differ by only 0.2 W. EDGE has the lower MAE in the other four pairs and the higher F_1 in all six. The microwave example demonstrates that the metrics can disagree under this protocol.

Table 2. **Power reconstruction and event recognition on shared test samples**

House	Appliance	Method	MAE, W	F_1 complete cycles [95 % CI]	F_1 Known onsets	Counts k / n / m	D_{TV} Complete cycles
2	Kettle	CO	44.7	0.170 [0.074, 0.271]	0.180	9 / 19 / 87	0.384
2	Kettle	EDGE	23.2	0.560 [0.421, 0.651]	0.571	14 / 19 / 31	0.285
2	Microwave	CO	31.2	0.015 [0.000, 0.041]	0.022	2 / 4 / 262	0.775
2	Microwave	EDGE	10.7	0.130 [0.000, 0.286]	0.136	3 / 4 / 42	0.571
4	Kettle	CO	26.3	0.032 [0.000, 0.079]	0.031	2 / 7 / 117	0.686
4	Kettle	EDGE	14.4	0.414 [0.200, 0.579]	0.457	6 / 7 / 22	0.636
4	Microwave	CO	18.9	0.079 [0.013, 0.168]	0.089	6 / 15 / 136	0.566
4	Microwave	EDGE	11.4	0.605 [0.381, 0.792]	0.652	13 / 15 / 28	0.362
5	Kettle	CO	29.2	0.077 [0.000, 0.250]	0.065	2 / 3 / 49	0.857
5	Kettle	EDGE	29.4	0.129 [0.000, 0.375]	0.151	2 / 3 / 28	0.679
5	Microwave	CO	52.1	0.013 [0.000, 0.058]	0.033	1 / 12 / 138	0.609
5	Microwave	EDGE	84.6	0.089 [0.000, 0.176]	0.097	4 / 12 / 78	0.538

Source: **compiled by the authors**

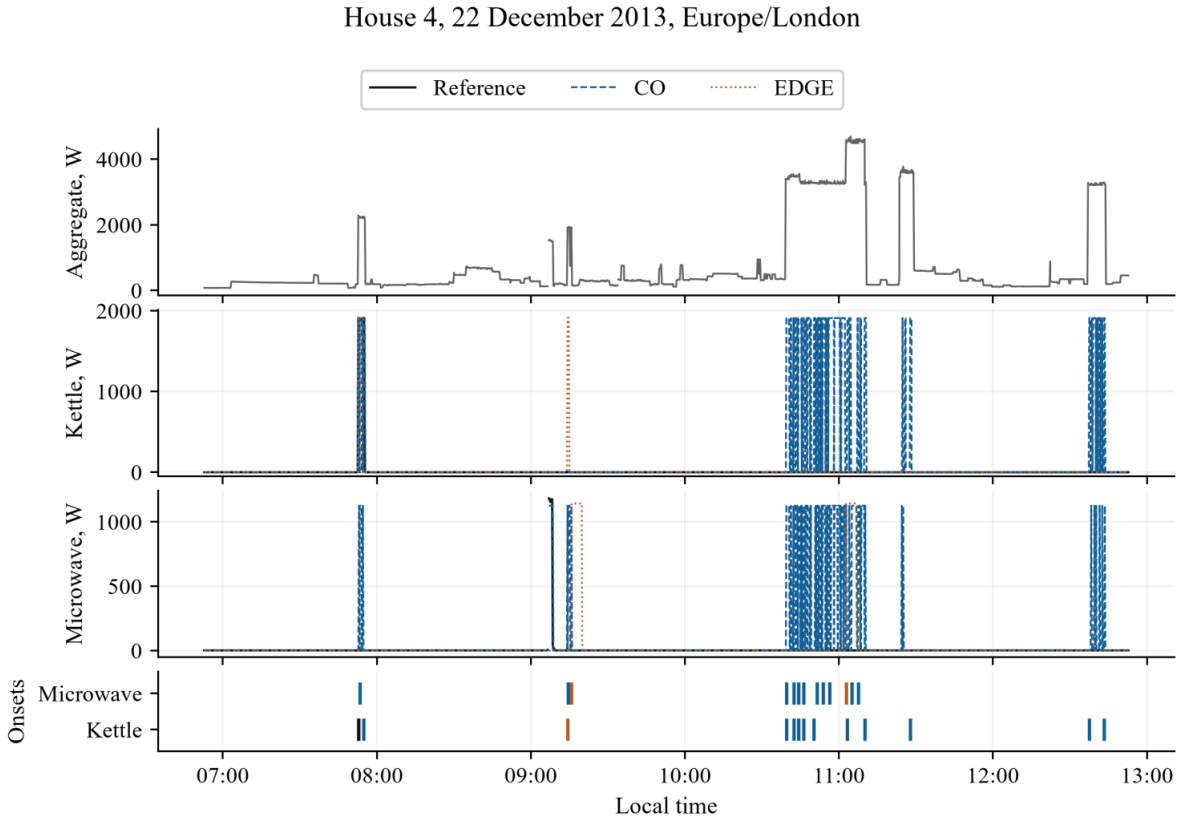


Fig. 2. **Aggregate power, reconstructed appliance power and confirmed kettle and microwave onsets in a test excerpt from House 4 on 22 December 2013**

Source: **compiled by the authors**

Overcounting and rapid state switching in CO. Because CO does not use the previous state, it generates many false detections. The kettle trace in Fig. 2 shows closely spaced, brief spurious transitions. For the microwave in House 2, CO estimates 262 events against 4 reference events (N_{est}/N_{ref} = 262/4), while EDGE estimates 42 events.

Effect of widening the matching tolerance. At 120 s, the absolute F_1 scores for CO increase slightly. EDGE retains its lead, and its own scores remain stable.

Fig. 3 compares the recovered daily onset profiles and event durations. For complete events, the EDGE onset profile is closer to the reference in all six pairs. For the kettle in House 2, the total variation distance is 0.285 for EDGE and 0.384 for CO. For the microwave in House 4, the corresponding distances are 0.362 and 0.566. Normalization hides differences in activation counts, so we examine the absolute counts alongside the profiles.

For example, the mean absolute error in the daily microwave onset count in House 2 is 2.22 for EDGE and 14.33 for CO.

The curves in Fig. 3 use complete operational events only. Pooling data from the three houses provides a descriptive comparison. Statistical uncertainty across independent facilities was not estimated from these curves.

Duration errors apply to matched pairs. For the kettle in House 4, CO has a mean duration error of 75.0 s on two matches, while EDGE has an error of 108.2 s on six matches. Their respective recalls are 0.286 and 0.857. For the microwave in House 4, the duration errors are 154.5 s on six matches for CO and 59.8 s on thirteen matches for EDGE.

The reference contains 60 complete cycles, 17 known onsets with unknown endings, and 520 fragments with unknown onsets. Several fragments can belong to one physical activation. In particular, the three complete kettle cycles in House 5 and the four complete microwave cycles in House 2 describe what the strict protocol could verify.

When the target is changed to all 77 known reference onsets, EDGE still has the higher F_1 in all six pairs (Table 2). This comparison retains the original segmentation and does not recover unknown starts.

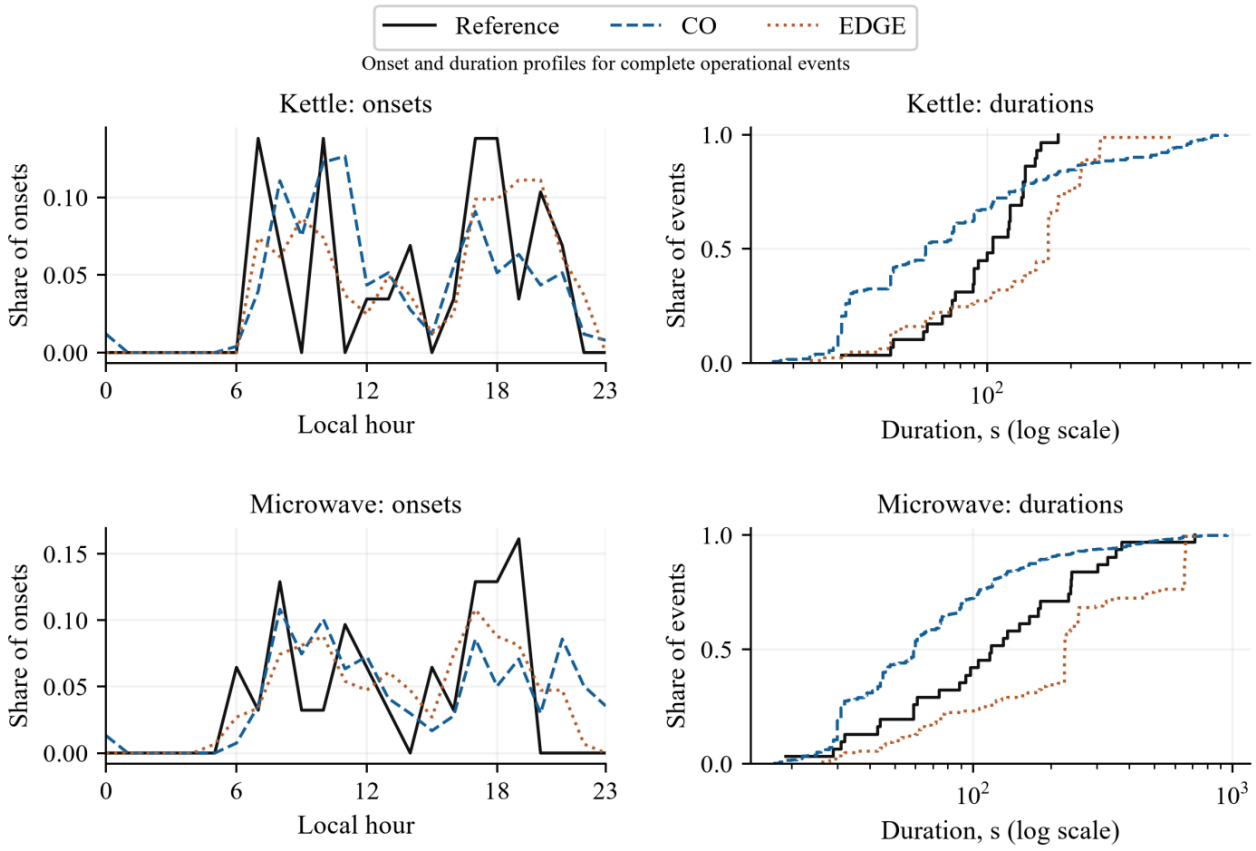


Fig. 3. **Hourly onset profiles and duration distributions for complete events pooled across the three test houses**

Source: **compiled by the authors**

With the quality mask fixed, the reference contains 60, 61, and 61 complete cycles at gap limits of 24, 60, and 120 s. Counts at these limits are House 2 kettle: 19/19/19; House 2 microwave: 4/4/4; House 4 kettle: 7/7/7; House 4 microwave: 15/15/15; House 5 kettle: 3/4/4; House 5 microwave: 12/12/12. A wider gap admits more uninterrupted labels across missing observations. This comparison measures sensitivity of the operational reference to segmentation.

Only one additional complete onset was recovered at a wider gap, for the House 5 kettle. It lasted 78 s under the relaxed rule, compared with a median of 100 s among the three onsets already complete at 24 s. This sample cannot establish whether censoring preferentially removes longer physical activations. Invalid samples and other segment boundaries still interrupt the event rule.

Changing both level thresholds by ±20% left complete-cycle F_1 unchanged in all six pairs. Varying persistence from 8 to 24 s changed the scores. For example, the House 5 kettle yielded 0.103 at 8 s, 0.129 at 16 s, and 0.154 at 24 s. The House 5 microwave yielded 0.088, 0.089, and 0.071, respectively. This one-factor check holds the saved EDGE power estimates and the reference labels fixed. It varies confirmation only.

The House 5 microwave cycle starting at 2013-12-07 08:51:36 UTC had no EDGE complete-cycle match within 60 s. Figure 4 shows the reference and estimated power around this onset.

Forecast accuracy and exceedance screening

The forecasts share 16,269, 10,734 and 10,539 test origins in Houses 2, 4 and 5, respectively. F_{EDGE} reduces MAE by 0.4, 1.4 and 4.8 W (Table 3). F_{CO} yields a smaller error than F_{base} only in House 4.

The paired MAE differences F_{EDGE} minus F_{base} and their day-block intervals were House 2, -0.40 W [-1.74, +1.26]; House 4, -1.40 W [-2.61, +0.03]; House 5, -4.81 W [-7.09, -2.74]. The interval is below zero in House 5 and includes zero in Houses 2 and 4. The comparisons are exploratory and conditional on the test records. F_{oracle} has a higher MAE than F_{base} in each house under the fixed linear specification. The results do not distinguish between the effects of regularization, model form, and variation in appliance use over time.

At origins followed by a known target-appliance onset, MAEs for F_{base}, F_{EDGE} and F_{oracle} were House 2: 65 origins, 3265.1/3247.1/3232.8 W; House 4: 59 origins, 1572.3/1581.4/1572.5 W; House 5: 17 origins, 1884.1/1946.4/1869.1 W. These small subsets give a descriptive check and exclude onsets that could not be verified.

Empirical test coverage ranged from 89.5 % to 91.4%. Underestimation in Table 3 is the mean of max(0, observed peak minus upper estimate) over all origins. In House 2 the F_{base} validation adjustment is 20.8 W, and in House 4 it is 219.8 W. These values describe the upper tail of signed validation errors. MAE averages errors of either sign on a different period, so it can be higher or lower than the adjustment [6], [25].

In Table 4, n is the number of origins with a measured exceedance. Recall and false-positive rate (FPR) are percentages. The alert recall varies from 16.7 % to 60.8 % across the tested settings, with little difference between feature variants. The results describe screening per forecast origin. They do not count distinct overload episodes or establish controller effectiveness.

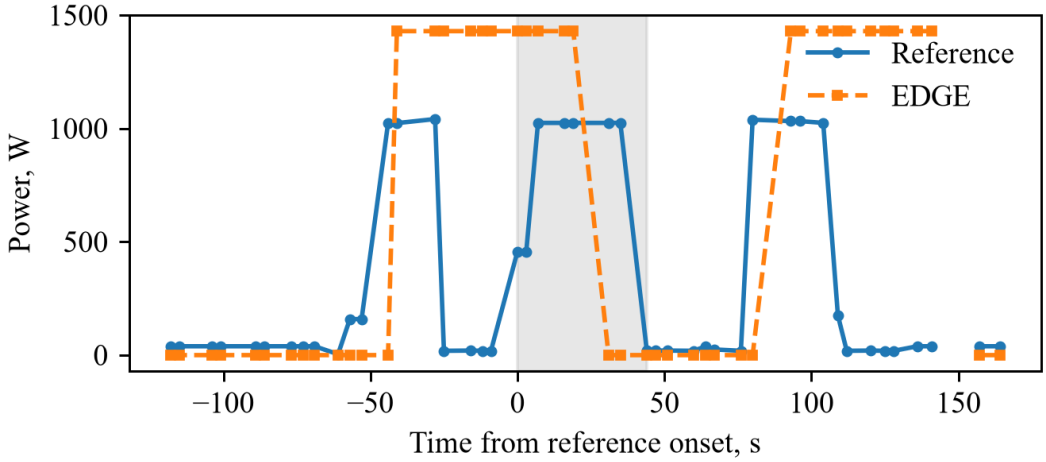


Fig. 4. **House 5 microwave cycle without an EDGE onset match**
Source: **compiled by the authors**

Table 3. **Forecasts of the aggregate maximum at shared test times**

House	Variant	MAE, W	Δ MAE vs F_{base}, %	Coverage, %	$q_{0.90}$, W	Mean under-estimation, W
2	F_{base}	169.1	0.00	90.4	20.8	86.8
2	F_{CO}	170.6	+0.86	90.6	26.2	86.4
2	F_{EDGE}	168.7	-0.24	90.2	21.0	86.2
2	F_{oracle}	170.9	+1.03	90.4	23.9	86.6
4	F_{base}	182.5	0.00	91.4	219.8	59.5
4	F_{CO}	181.6	-0.51	91.3	219.0	59.5
4	F_{EDGE}	181.1	-0.77	91.3	219.5	60.0
4	F_{oracle}	183.2	+0.39	91.3	220.4	59.9
5	F_{base}	139.2	0.00	89.8	59.7	48.4
5	F_{CO}	140.2	+0.74	89.6	59.6	48.8
5	F_{EDGE}	134.4	-3.46	89.5	58.9	49.0
5	F_{oracle}	143.7	+3.24	90.0	66.4	48.5

Source: **compiled by the authors**

Table 4. **Exceedance screening at test forecast origins**

House / kW	n	F_{base} recall / FPR	F_{EDGE} recall / FPR
2 / 3	466	57.9 / 0.07	58.4 / 0.08
2 / 4	184	54.3 / 0.04	54.3 / 0.05
2 / 5	172	52.3 / 0.02	52.3 / 0.01
4 / 3	314	60.8 / 0.06	60.8 / 0.05
4 / 4	25	40.0 / 0.05	40.0 / 0.06
4 / 5	6	16.7 / 0.00	16.7 / 0.00
5 / 3	143	31.5 / 0.32	31.5 / 0.34
5 / 4	49	42.9 / 0.10	40.8 / 0.10
5 / 5	24	41.7 / 0.08	41.7 / 0.07

Source: **compiled by the authors**

For example, at 2013-12-22 10:39:48 UTC in House 4, F_{EDGE} gives an upper estimate of 3609 W and headroom of 1391 W at the illustrative 5 kW limit. A planned operation drawing up to 1 kW over that five-minute horizon would pass this screen. A 1.5 kW operation would be deferred to a later origin. The observed peak was 3557 W before any hypothetical addition. This calculation assumes additive load and an operation contained within the horizon.

DISCUSSION OF RESULTS

The proposed method integrates preprocessing, rolling statistical descriptors, causal disaggregation, event validation and peak-load forecasting into a single data mining pipeline for proactive energy consumption. CO, EDGE and ridge regression provide the implementations used to evaluate this pipeline. Their comparison on shared records makes the effects of the reconstructed appliance information observable at both the event and forecast stages. The experiment does not isolate the contribution of every stage.

Both baselines need labeled appliance channels for local calibration. After calibration, NILM inference uses the aggregate meter. The recorded British household loads from 2013-2014 provide a limited basis for extrapolation to commercial or industrial buildings, where appliance signatures and concurrent transitions can differ. A centered-window model such as RAPC-Net [16] also uses future samples, which changes its suitability for a forecast issued at the current time.

The data use a nominal cadence of about 8 s, with irregular sampling. The confirmation rule waits at least 16 s from the first qualifying sample and can wait longer for the next observation. A confirmed state can enter at the next retained one-minute forecast origin when the data are valid. EDGE compares each observed transition with the fitted signatures of the two target appliances. CO compares each sample against up to 512 precomputed sums, and ridge inference is linear in

the number of features. Total runtime and communication latency were not measured.

CONCLUSIONS

The proposed data mining method for non-intrusive load monitoring integrates five stages in a causal pipeline. Preprocessing and statistical feature engineering with calendar context are followed by disaggregation with confirmation rules, one-to-one validation of operational events and forecasting of available power. REFIT household records provide an experimental evaluation of the method.

EDGE achieved higher complete-cycle F_1 than CO in all six evaluated house–appliance pairs and lower hourly-profile distance in each pair. For the House 4 microwave, F_1 increased from 0.079 to 0.605. The House 5 microwave illustrates why power MAE and event recognition should be assessed separately.

Reference completeness depended on the allowed observation gap. Reporting the complete cycles and the partially observed activations separately made clear how much of the evidence supported direct comparison of onset counts and durations.

With ridge regression, adding EDGE features reduced peak-forecast MAE by 0.4, 1.4 and 4.8 W, or 0.24 %, 0.77 % and 3.44 %, in Houses 2, 4 and 5. The day-block intervals included zero in two houses and were below zero in House 5. These improvements establish a local benefit under the fixed forecast protocol.

The validation error quantile and the resulting power headroom supplied a forecast-based input for proactive resource allocation. Empirical test coverage was 89.5-91.4 %. The decision example connects this estimate to screening a planned operation against an assumed power limit.

Conflict of interest. The authors declare that they have no conflict of interest

Funding. The authors funded the study without external support

Data availability. The original records are available through REFIT [20] and the associated Zenodo record (https://doi.org/10.5281/zenodo.5063428). The code, configurations, manifests, forecast series and source tables are stored in a local reproduction package. No public repository is available for that package

Use of artificial intelligence. AI tools were used to translate the manuscript from Ukrainian into English, format the manuscript, and prepare figures and tables from the study data. The authors take responsibility for the content of the article

REFERENCES

1. Azizi, E., Beheshti, M. T. H. & Bolouki, S. “Event matching classification method for non-intrusive load monitoring”. *Sustainability*. 2021; 13 (2): 693. DOI: https://doi.org/10.3390/su13020693.

2. Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A. & Srivastava, M. “NILMTK: An open source toolkit for non-intrusive load monitoring”. In *Proceedings of the* 5th *ACM International Conference on Future Energy Systems*. 2014. p. 265–276. DOI: https://doi.org/10.1145/2602044.2602051.

3. Batra, N., Kukunuri, R., Pandey, A., Malakar, R., Kumar, R., Krystalakos, O., Zhong, M., Meira, P. & Parson, O. “Towards reproducible state-of-the-art energy disaggregation”. In: *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 2019. p. 193–202. DOI: https://doi.org/10.1145/3360322.3360844.

4. D’Incecco, M., Squartini, S. & Zhong, M. “Transfer learning for non-intrusive load monitoring”. *IEEE Transactions on Smart Grid*. 2020; 11 (2): 1419–1429. DOI: https://doi.org/10.1109/TSG.2019.2938068

5. Gao, D., Peng, B., Zhi, Y., Zhang, X. & Yang, X. “NILM-Flex: Quantifying residential demand-side flexibility potential via Non-intrusive load monitoring and behavior-aware appliance modeling”. *Applied Energy*. 2026; 422: 128315. DOI: https://doi.org/10.1016/j.apenergy.2026.128315.

6. Gibbs, I. & Candès, E. “Adaptive conformal inference under distribution shift”. *Advances in Neural Information Processing Systems*. 2021; 34: 1660–1672. – Available from: https://papers.nips.cc/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html.

7. Hart, G. W. “Nonintrusive appliance load monitoring”. *Proceedings of the IEEE*. 1992; 80 (12): 1870–1891. DOI: https://doi.org/10.1109/5.192069.

8. Hoerl, A. E. & Kennard, R. W. “Ridge regression: Biased estimation for nonorthogonal problems”. *Technometrics*. 1970; 12 (1): 55–67. DOI: https://doi.org/10.1080/00401706.1970.10488634.

9. Kelly, J. & Knottenbelt, W. “Neural NILM: Deep neural networks applied to energy disaggregation”. *In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. 2015. p. 55–64. DOI: https://doi.org/10.1145/2821650.2821672.

10. Kelly, J. & Knottenbelt, W. “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes”. *Scientific Data.* 2015; 2: 150007. DOI: https://doi.org/10.1038/sdata.2015.7.

11. Klemenjak, C., Reinhardt, A., Pereira, L., Makonin, S., Bergés, M. & Elmenreich, W. “Electricity consumption data sets: Pitfalls and opportunities”. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 2019. p. 159–162. DOI: https://doi.org/10.1145/3360322.3360867.

12. Klemenjak, C., Kovatsch, C., Herold, M. & Elmenreich, W. “A synthetic energy dataset for non-intrusive load monitoring in households”. *Scientific Data.* 2020; 7: 108. DOI: https://doi.org/10.1038/s41597-020-0434-6.

13. Klemenjak, C., Makonin, S. & Elmenreich, W. “Towards comparability in non-intrusive load monitoring: On data and performance evaluation”. In *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. 2020. p. 1–5. DOI: https://doi.org/10.1109/ISGT45199.2020.9087706.

14. Kolter, J. Z. & Jaakkola, T. S. “Approximate inference in additive factorial HMMs with application to energy disaggregation”. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. 2012; 22:1472–1482. – Available from: https://dspace.mit.edu/entities/publication/d7e1d6d8-5c47-492f-a706-2c1034f83a41.

15. Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y. & Zhang, Y. “Short-term residential load forecasting based on LSTM recurrent neural network”. *IEEE Transactions on Smart Grid.* 2019; 10 (1): 841–851. DOI: https://doi.org/10.1109/TSG.2017.2753802.

16. Li, K. & Sun, C. “RAPC-Net: Residual-aware physical consistency network for multi-appliance NILM”. *Applied Sciences.* 2026; 16 (17): 8866. DOI: https://doi.org/10.3390/app16178866.

17. Li, Y., Wang, H., Yang, Z., Chen, Z. & Yang, J. “Two-stage ultra-short-term load forecasting model of household appliances based on non-intrusive load disaggregation”. *Transactions of China Electrotechnical Society*. 2024; 39 (11): 3379–3391 (in Chinese), https://www.scopus.com/pages/publications/85197216106. DOI: https://doi.org/10.19595/j.cnki.1000-6753.tces.230554.

18. Lucas, A., Jansen, L., Andreadou, N., Kotsakis, E. & Masera, M. “Load flexibility forecast for DR using non-intrusive load monitoring in the residential sector”. *Energies*. 2019; 12 (14): 2725. DOI: https://doi.org/10.3390/en12142725.

19. Murray, D., L. Stankovic, and V. Stankovic. “An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study”. *Scientific Data.* 2017; 4: 160122. DOI: https://doi.org/10.1038/sdata.2016.122.

20. Murray, D., Stankovic, L. & Stankovic, V. “REFIT: Electrical Load Measurements (Cleaned)”. *Dataset. University of Strathclyde*, 2016. DOI: https://doi.org/10.15129/9ab14b0e-19ac-4279-938f-27f643078cec.

21. Precioso, D. & Gómez-Ullate, D. “Thresholding methods in non-intrusive load monitoring”. *The Journal of Supercomputing*. 2023; 79 (13): 14039–14062. DOI: https://doi.org/10.1007/s11227-023-05149-8.

22. Rehman, A. U., Lie, T. T., Valles, B. & Tito S. R. “Event-detection algorithms for low sampling nonintrusive load monitoring systems based on low complexity statistical features”. *IEEE Transactions on Instrumentation and Measurement*. 2020; 69 (3): 751–759. DOI: https://doi.org/10.1109/TIM.2019.2904351.

23. Rodriguez-Navarro, C., Portillo, F., Robalo, I. & Alcayde, A. “Evaluation of traditional and data-driven algorithms for energy disaggregation under sampling and filtering conditions”. *Inventions*. 2025; 10 (3): 43. DOI: https://doi.org/10.3390/inventions10030043.

24. Welikala, S., Dinesh, C., Ekanayake, M. P. B., Godaliyadda, R. I. & Ekanayake, J. “Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting”. *IEEE Transactions on Smart Grid*. 2019; 10 (1): 448–461. DOI: https://doi.org/10.1109/TSG.2017.2743760.

25. Xu, C. & Xie, Y. “Conformal prediction interval for dynamic time-series”. In *Proceedings of the 38th International Conference on Machine Learning*, 2021; 139: 11559–11569. – Available from: https://proceedings.mlr.press/v139/xu21h.html.

26. Zhang, C., Zhong, M. Wang, Z., Goddard, N. & Sutton, C. “Sequence-to-point learning with neural networks for non-intrusive load monitoring”. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018; 32 (1). DOI: https://doi.org/10.1609/aaai.v32i1.11873.

27. Zhao, B., He, K., Stankovic, L., & Stankovic, V. “Improving event-based non-intrusive load monitoring using graph signal processing”. *IEEE Access*. 2018; 6: 53944–53959. DOI: https://doi.org/10.1109/ACCESS.2018.2871343.

Received 29.08.2026
Received after revision 26 10.2026
Accepted 30.10.2026

DOI: https://doi.org/10.15276/aait.09.2026.36
УДК 004.62:621.31

Метод інтелектуального аналізу даних неінтрузивного моніторингу електричних навантажень у системах проактивного енергоспоживання смарт-об’єктів

Засідько Андрій Ярославович[1)]
ORCID: https://orcid.org/0000-0001-5197-0435; andzasidko@op.edu.ua. Scopus Author ID: 57792636800
Арсірій Олена Олександрівна[1)]
ORCID: https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900
[1)] Національний університет «Одеська політехніка», пр. Шевченка, 1, Одеса, 65044, Україна

АНОТАЦІЯ

Для завчасного виділення потужності на заплановану операцію смарт-об’єкту потрібна оцінка майбутнього пікового навантаження. У роботі розроблено метод інтелектуального аналізу даних неінтрузивного моніторингу електричних навантажень у системах проактивного енергоспоживання смарт-об’єктів, який об’єднує попередню обробку даних, виділення операційних подій та прогностичне моделювання в єдиний архітектурний конвеєр. Метод забезпечує перетворення «сирих» часових рядів агрегованого споживання на прикладі датасету REFIT у комплекс інформаційних артефактів, а саме дискретні операційні події та їх атрибути (часи початку, тривалості, добові профілі використання), квантифіковані оцінки точності розпізнавання та прогнозний базис для управління у вигляді квантильної оцінки запасу потужності в контурі проактивного енергоспоживання. Проведено порівняльний аналіз алгоритмів двостанової комбінаторної оптимізації та причинного інженерного детектора змін потужності. Експериментально показано можливість розбіжності між ранжуванням моделей за похибкою відновлення неперервної потужності та якістю розпізнавання дискретних подій. У межах фінального етапу методу інтегровано виділені приладові ознаки в регуляризовану лінійну регресію, сформовано валідаційні квантилі похибки та розраховано квантильну оцінку запасу потужності. Приладові ознаки детектора зменшили середню абсолютну похибку прогнозу п’ятихвилинного максимуму на нуль цілих чотири десятих, одну цілу чотири десятих та чотири цілих вісім десятих вата у трьох будинках. Детектор мав вищий F_1 розпізнавання подій у всіх шести парах «будинок–прилад». Отримана оцінка дає системі проактивного керування змогу визначити доступну потужність на горизонті прогнозування.

Ключові слова: інтелектуальний аналіз даних; неінтрузивний моніторинг навантажень; дезагрегація часових рядів; моделювання операційних подій; прогнозування часових рядів; проактивне енергоспоживання

ABOUT THE AUTHORS



Andrii Ya. Zasidko - PhD student, assistant lecturer, Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-5197-0435; andzasidko@op.edu.ua. Scopus Author ID: 57792636800
Research field: Computer science

Засідько Андрій Ярославович - аспірант, асистент кафедри Інформаційних систем. Національний університет «Одеська політехніка», пр. Шевченка, 1, Одеса, 65044, Україна
Галузь досліджень: Комп'ютерні науки



Olena O. Arsirii - Doctor of Engineering Sciences, Professo, Head of Department of Information Systems. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-8130-9613; e.arsiriy@gmail.com. Scopus Author ID: 54419480900***Research field:*** ***Research field:*** Information technology, artificial intelligence, decision support systems, machine learning, and neural networks

Арсірій Олена Олександрівна - доктор технічних наук, професор, завідувач кафедри Інформаційних систем. Національний університет «Одеська політехніка», просп. Шевченка, 1, Одеса, 65044, Україна
Галузь досліджень: Інформаційні технології, штучний інтелект, системи підтримки рішень, машинне навчання та нейронні мережі