Forecast-driven change point detection in electricity consumption time series

***Mathias de Schietere de Lophem*[a], *Rosana Veroneze*[b], *Axel Legay*[c], *Kim Mens*[d], *Hélène Verhaeghe*[e]**

[a] *UCLouvain, Louvain-la-Neuve, Belgium, mathias.deschietere@uclouvain.be CA*
[b] *UCLouvain, Louvain-la-Neuve, Belgium, rosana.veroneze@uclouvain.be*
[c] *Legay Consulting, Belgium, legay@legayconsulting.be*
[d] *UCLouvain, Louvain-la-Neuve, Belgium, kim.mens@uclouvain.be*
[e] *UCLouvain, Louvain-la-Neuve, Belgium, helene.verhaeghe@uclouvain.be*

Abstract:
Detecting persistent structural changes in electricity consumption is essential for energy management in buildings, smart grids, and decentralized systems, because regime shifts in occupancy, appliances, technologies, or usage habits degrade forecasting performance and operational decisions. Classical change-point detection (CPD) applied directly to raw load is often unreliable since electricity series are multi-scale, highly seasonal, weather-sensitive, and behaviour-driven, predictable variability can be mistaken for structural change, while gradual genuine shifts can be missed.
To address this limitation, we explored a forecasting-based CPD framework for electricity-consumption analysis. Expected demand is first estimated with statistical and machine-learning forecasting models using temporal and contextual drivers, then CPD is applied to residuals (observed minus predicted load). This shifts detection from raw variability to unexplained deviations, aiming to reduce spurious alarms and improve sensitivity to meaningful persistent changes.
An important side contribution is a controlled benchmark based on synthetic electricity datasets with ground-truth change-point labels, designed to preserve realistic seasonal and operational behaviour. Our evaluation covers six datasets (three EnergyPlus-based and three LUCID), fourteen forecasting models, and thirteen CPD methods, under identical settings for raw and residual pipelines, using tolerance windows $\delta \in \{1, 3, 7\}$ and metrics including F1 Score, G-mean, and Covering.
Results show a conditional rather than universal advantage for forecast-driven CPD. At $\delta = 7$, global counts yield 285 residual wins, 466 raw wins, and 2525 ties across F1 Score, G-mean, and Covering metrics. Whereas residual gains concentrate in specific algorithms (notably PELT, Binseg, Window, BottomUp, and KernelCPD), pointwise metrics still favour raw CPD in aggregate, while Covering is slightly favours forecast-driven CPD (163 residual vs. 152 raw wins), indicating potential segmentation gains in selected settings. Overall, forecast-driven CPD should be deployed adaptively, not as a default preprocessing step.

Keywords:
Change point detection, Electricity consumption time series, Structural change analysis, Synthetic datasets, Residual analysis.

1. Introduction

Transition toward low-carbon energy systems has increased the use of fine-resolution electricity data in buildings, energy communities, and decentralized power systems. These data support forecasting, fault detection, performance assessment, and operational control. Their value, however, depends on the ability to distinguish routine variability from persistent changes in electricity consumption.
Detecting structural changes in consumption patterns is a challenge. These changes can result from occupancy shifts, schedule modifications, equipment replacement, electrification of end uses, or

control-strategy updates. Because forecasting and monitoring models are trained on historical behaviour, regime changes degrade model accuracy and decision quality. Energy-forecasting studies confirm that demand is strongly shaped by seasonality, calendar effects, and exogenous drivers, and that abrupt societal or operational shifts can quickly invalidate previously accurate models [14, 26].

Change point detection (CPD) is a framework for identifying shifts in time series [19, 39]. In energy applications, CPD has been used in market analysis, grid monitoring, and building analytics, but direct application to raw electricity load is difficult. Load series are multi-scale, weather-sensitive, and structured by daily, weekly, and seasonal routines, so changes in mean or variance may reflect predictable variation rather than meaningful behavioural change. Operational changes and non-routine events are hard to isolate without a baseline model of expected demand [12,38], and direct monitoring of raw observations can yield many false alarms under changing conditions [8, 22, 23, 36].

These limitations motivate a forecasting-based view of structural change detection: expected demand is estimated from temporal and exogenous covariates, then CPD is applied to residuals (i.e., unexplained consumption). This logic appears in measurement and verification [12, 38], abnormal-consumption detection [20, 23], and drift-aware smart-grid monitoring [8, 13, 25]. However, prior work mostly targets anomaly screening, fault detection, or forecast adaptation rather than CPD on residuals for persistent structural-change detection.

Another challenge is that public electricity datasets rarely include reliable annotations for the timing and causes of structural changes, which limits rigorous CPD comparison. Partial solutions include annotated anomaly datasets [11], automatic labeling for abnormal consumption [10], and synthetic anomaly injection [42]. Although synthetic energy time-series generation has expanded rapidly, most datasets target privacy, scenario generation, or model training rather than structural-change benchmarking with known ground truth [43].

The main contributions of this paper are: 1) A forecasting-based framework for structural change detection in electricity consumption time series. The load is estimated using regressions, and CPD is performed on the residual series. 2) Synthetic datasets of consumption with labelled change points, enabling controlled evaluation, designed to preserve realistic seasonality, exogenous sensitivity, and operational variability. 3) A systematic benchmarking setting for comparing CPD performance under conditions representative of building and household electricity use.

The scope of this work is limited to univariate electricity consumption time series at building or household scale. The events are persistent behavioural or operational shifts rather than isolated outliers.

2. Methodology and Experimental Protocol

This section presents the workflow used in this study, from forecasting and residual construction to CPD, evaluation, and raw-versus-residual comparison.

Let $\{y_t\}_{t=1}^{T}$, with t the time index, denote a univariate electricity-consumption time series aggregated over a fixed interval (e.g., hourly or daily). Consumption is driven by daily/weekly patterns, seasonality, weather, and behavioural variability. The observed signal is therefore decomposed into predictable and residual components: $y_t = f(t, \mathbf{x}_t) + \varepsilon_t$. The expected electricity demand is $f(t, \mathbf{x}_t)$, a function of time and contextual variables $\mathbf{x}_t$, and the residual, capturing unexplained deviations, is ε_t. The contextual variables $\mathbf{x}_t$ combine lag features, trend components, seasonality terms, calendar indicators, weather variables, and rolling statistics from past consumption.

In this context, a *structural change* is a persistent modification in underlying consumption behaviour that alters the data-generating process over an extended period. Examples include occupancy shifts, acquisition of new appliances, or long-term habit changes. These differ from short-term fluctuations, noise, and predictable seasonal variation inherent to electricity-consumption data [1, 38].

CPD identifies time instants at which time-series' statistical properties change. Applied directly to raw electricity consumption $\{y_t\}$, predictable effects such as seasonality and weather-driven variability often violate stationarity assumptions used by many methods, leading to frequent false detections [44]. To address this issue, change detection is defined on residuals $\{\varepsilon_t\}$ rather than raw data. Expected consumption $f(t, \mathbf{x}_t)$ is estimated with forecasting models, and CPD is applied to residuals so that detection focuses on deviations unexplained by known drivers and regular patterns. Detected change points are thus more likely to represent genuine persistent behavioural or structural shifts. The proposed framework has two stages:

1. Forecast expected electricity consumption. Models are trained with the engineered features described above. Hyperparameters are selected by grid search. Forecasting quality is assessed with RMSE (emphasizes large errors), MAE (absolute average deviation), and MAPE (scale-normalized but unstable near zero) [16].
2. Apply CPD to both raw series $\{y_t\}$ and residual series $\{\varepsilon_t\}$ ($\varepsilon_t = y_t - \hat{y}_t$) under identical settings. CPD quality is measured using the *F1 Score* (precision–recall balance), the *G-mean* (balance between true positive and false positive rates under class imbalance), and *Covering* (agreement between estimated and true segmentations) [44], as they jointly capture detection accuracy, robustness to class imbalance, and segmentation agreement.

The central hypothesis is that removing predictable effects prior to CPD can reduce spurious detections due to seasonality and exogenous variability.

2.1. Experimental Setting

Experiments are conducted on six synthetic datasets. For each dataset, the first two years are used for model fitting and parameter selection, and the third year is reserved for out-of-sample CPD evaluation. This chronological split prevents temporal leakage.

Fourteen forecasting methods are tested: autoregressive integrated moving average (ARIMA) [4], seasonal autoregressive integrated moving average with exogenous regressors (SARIMAX) [33], linear regression (LR) [24], lasso [37], stochastic gradient descent regression (SGD) [3], k-nearest neighbors regression (KNN) [7], least-squares support vector regression (LSVR) [35], support vector regression (SVR) [34], multilayer perceptron (MLP) [31], extreme gradient boosting (XGB) [6], and four hybrid ensembles combining two complementary forecasting approaches (such as LR+XGB).

Thirteen CPD methods spanning multiple families are tested: pruned exact linear time (PELT) [18], binary segmentation (Binseg) [32], bottom-up segmentation (BottomUp) [17], sliding-window segmentation (Window) [2], and kernel change-point detection (KernelCPD) [5] from `ruptures` [40, 41]; cumulative sum (CUSUM) [27]; exponentially weighted moving average (EWMA) [28]; and two-sample test statistic (TwoSample) [29] from `ocpdet` [45]; sequential and batch change detection within the `cpm` framework [30], which relies on two-sample test statistics, including single change-point detection in a sequential setting (cpm1S), single change-point detection in a batch setting (cpm1B), and multiple change-point detection in a sequential setting (cpmMS); and standard binary segmentation (sbs) and wild binary segmentation (wbs) from the `wbs` framework [9]. Some entries should be interpreted as implementation-level variants rather than fully independent approaches. In particular, standard binary segmentation appears in `ruptures` and `wbs` libraries, and two-sample–based methods are implemented in the `ocpdet` and the `cpm` libraries. We report them separately, as implementation choices can lead to different behaviours under identical datasets and evaluation settings.

The main CPD evaluation parameters are: minimum segment length: 28 days, enforcing month-scale regimes, filtering short-lived fluctuations, and limiting over-segmentation of daily noise; sliding window: 28 days, giving local tests enough context across multiple weekly cycles, improving robustness to day of the week effects while preserving sensitivity to sustained changes; tolerance for matching detected and true change points: $\delta \in \{1, 3, 7\}$ days; a detection is counted as a true positive if it occurs within $\pm\delta$ days of a ground-truth (true) change point.

2.2. Comparison Procedure

For each metric, raw and residual configurations are compared pairwise for each dataset–CPD–forecast combination. Outcomes are labelled as residual better, raw better, or tie. Global summaries report: number of residual wins (residual CPD better on that metric); number of raw wins (raw CPD better); number of ties (equal performance); residual win rate (percentage of non-tie comparisons where residual CPD wins). This protocol provides a scale-agnostic and transparent way to determine whether forecast-driven CPD yields practical gains, and under which conditions.

2.3. Datasets

Objective evaluation of CPD for electricity consumption is limited by the lack of data with reliable ground-truth annotations of behavioural or structural changes. Smart-meter data are increasingly available, but often do not document when, how, or why consumption changes. Determining whether a detected change reflects genuine behavioural modification, long-term operational shift, or expected seasonal variability is thus often impossible. This limits assessment on real-world data alone.

For this reason, our study relies on synthetic electricity-consumption datasets. Synthetic generation provides full control over the timing, magnitude, and persistence of structural changes and yields unambiguous ground truth. This controlled setting enables objective, reproducible, and systematic CPD evaluation under strong seasonality and intrinsic variability [40].

2.3.1. EnergyPlus as a Data Source

Synthetic electricity-consumption time series are generated with EnergyPlus[1], a whole-building simulation engine developed by the U.S. Department of Energy and widely used in building engineering [15]. Building thermodynamics is modelled through geometry, materials, envelope properties, HVAC systems, internal gains, occupancy, equipment usage, and local weather. Electricity demand is derived from time-resolved energy-balance equations rather than ad hoc statistical assumptions, reproducing realistic daily/weekly/seasonal patterns and weather-driven variability. It is used here to generate controlled electricity-consumption scenarios for data-driven CPD. This preserves physical realism while retaining control over operational conditions and behavioural changes.

Source Simulation Profiles Dataset construction starts from multiple EnergyPlus simulations covering different building archetypes and operating scenarios. Simulations follow ASHRAE 90.1 standard reference models and include three representative buildings. Buildings are simulated under the Denver–Aurora–Buckley climate to keep weather conditions consistent. For each building type, multiple variants are generated to represent distinct operational regimes (baseline, occupancy-schedule changes, equipment-use changes, weekend patterns, night-shift operation, and PV-enabled operation). Each variant defines a stable regime used as a building block for multi-year synthetic series.

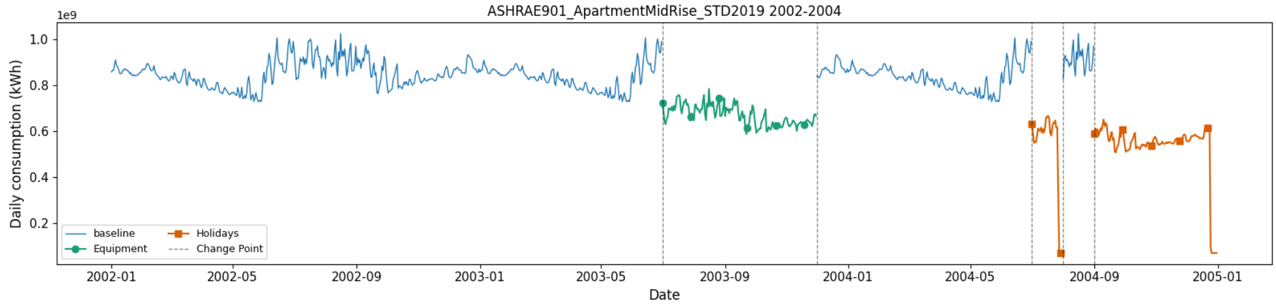
Synthetic Time Series Construction To generate long-horizon electricity-consumption series with controlled structural changes, each hourly EnergyPlus profile is transformed into a calendar-aligned lookup table indexed by month, day, and hour. This representation preserves daily, weekly, and seasonal structure, as well as weather-driven variability, when regimes are recombined.

A three-year synthetic dataset is then built for each building (Fig. 1). Year one is a pure baseline, providing a stable reference without behavioural changes. Years two and three are random concatenations of segments drawn from different regimes. Segment lengths range from one to six months, and transitions occur only at segment boundaries. This design creates realistic long-term consumption changes while keeping structural changes persistent and unambiguous.

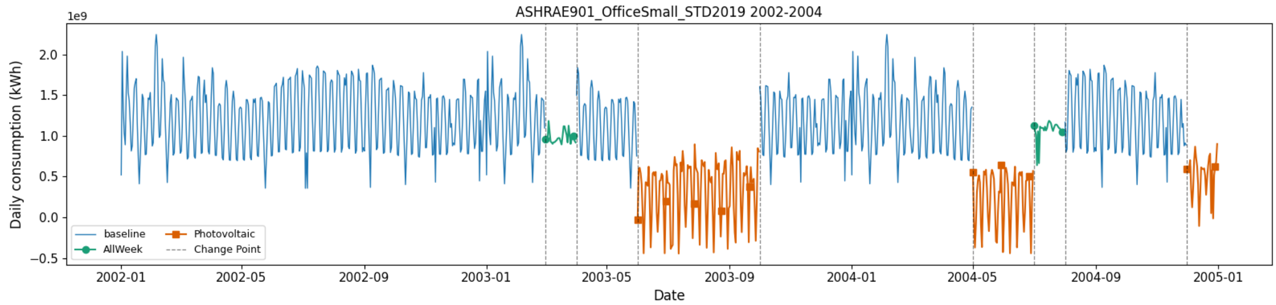
Transition rules ensure temporal consistency: 1) consecutive segments cannot use the same regime; and 2) the baseline regime is always enforced as the first segment of years two and three, limiting

[1] `https://energyplus.net`

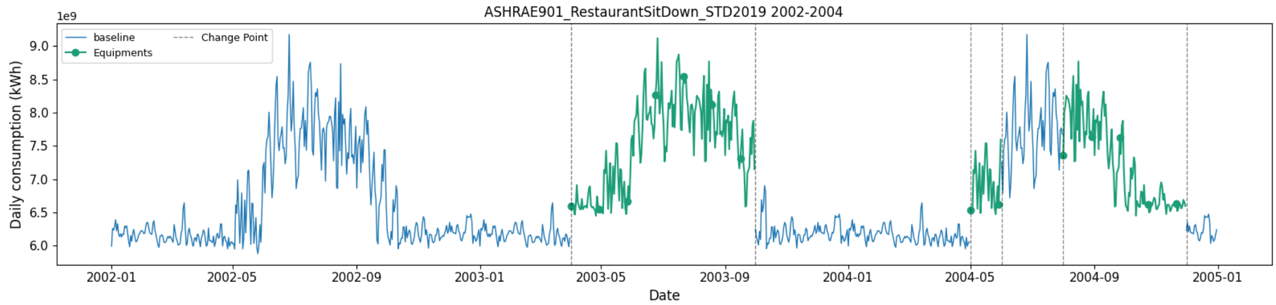
unrealistic multi-year drift and anchoring each year to a consistent start.



(a) Synthetic ApartmentMidRise 2002–2004



(b) Synthetic OfficeSmall 2002–2004



(c) Synthetic RestaurantSitDown 2002–2004

Figure 1: EnergyPlus datasets

Scenario Descriptions for Each Building Archetype Each building yields a set of distinct operational regimes derived from modified EnergyPlus schedules. These profiles represent realistic long-term behavioural/operational changes and serve as elementary regimes for multi-year synthesis. For each archetype, baseline and alternative regimes are summarized below.

Apartment Mid-Rise (Fig. 1a): The mid-rise apartment dataset contains three operational regimes:

- **Baseline.** Standard ASHRAE 90.1 residential schedules for a small household, with typical morning/evening peaks, lower midday usage, and heating/cooling-driven seasonality.
- **Equipment.** Sustained increase in household equipment usage relative to baseline. Appliance-related demand remains higher throughout the day while occupancy and weather are unchanged, yielding a higher non-HVAC base load and stronger activity peaks.
- **Holidays.** Long periods of near-unoccupancy (e.g., vacations). Lighting, equipment, occupancy, and domestic hot-water usage are strongly reduced, and HVAC runs in setback mode. Daily consumption drops substantially, with flatter profiles and very low activity peaks.

Small Office (Fig. 1b): The small office dataset also features three distinct operational profiles:

- **Baseline.** Standard five-day office operation with daytime occupancy. Electricity demand is dominated by HVAC, lighting, and office equipment, with low night/weekend usage.

- **Extended Operation (AllWeek).** Transition from weekday-only operation to seven-day activity. Weekend occupancy and internal gains become weekday-like, increasing weekly demand and reducing weekday/weekend contrast.
- **Photovoltaic (PV-assisted operation).** Building configuration with on-site photovoltaic generation. Occupancy and internal loads remain close to baseline, but net demand is reduced during daylight due to PV production, creating sharper afternoon drops.

Sit-Down Restaurant (Fig. 1c): Three operational regimes representing different business modes:

- **Baseline.** Standard ASHRAE restaurant with lunch and dinner service. The load profile has two peaks linked to kitchen equipment, ventilation, HVAC, and lighting during meals.
- **Increased Appliances (expanded kitchen).** Restaurant with a larger/more heavily equipped kitchen. Additional refrigeration and cooking appliances increase baseline demand; refrigeration stays elevated throughout the day and cooking peaks become stronger due to higher equipment density. Systematically higher consumption, especially during preparation and service.

Ground-Truth Change Point Labelling and Aggregation Because segmentation is fully controlled, the timing and type of each structural change are known. A change point is defined at the start of each new segment and corresponds to a persistent shift in operational regime.

Hourly series are aggregated to daily electricity consumption before CPD. This aligns with common monitoring practice and avoids focusing on high-frequency variability (diurnal cycles, occupancy micro-patterns, weather fluctuations) that can obscure structural behavioural shifts at hourly scale. Daily aggregation filters these effects and isolates long-term dynamics relevant to change detection.

2.3.2. LUCID Synthetic Residential Profiles

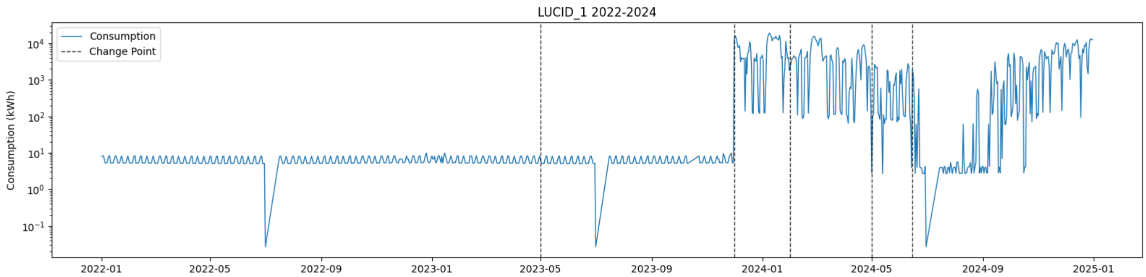
This study also uses synthetic residential electricity-consumption profiles generated in the LUCID laboratory (Belgium) [21]. STELLA[2], a system-dynamics language for complex dynamic systems, was used to generate them. The resulting models, incorporating the energy flow, exogenous inputs, and probabilistic behavioural rules, produce daily electricity-consumption trajectories consistent with the scenario. Each dataset contains three years of daily electricity consumption (2022–2024) of fictive households located in Liège (Belgium). Weather inputs come from the Open-Meteo Historical Forecast API for the corresponding coordinates/elevation at daily resolution.

Three residential scenarios are provided, corresponding to households with different initial states. For each, year one is a stable baseline, while structural changes are introduced progressively in years two and three. Changes represent realistic long-term shifts in household composition, appliance ownership, and energy technologies. A difference with EnergyPlus is that changes are *cumulative*: modifications can overlap and accumulate rather than replace one regime with another. The datasets are:

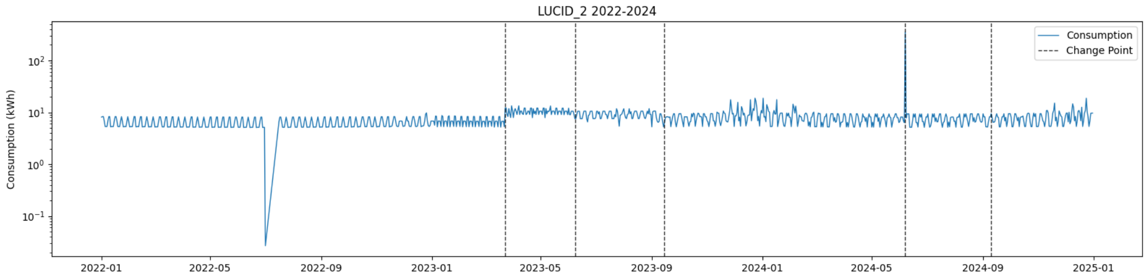
- **LUCID_1 (Fig. 2a)**: A 3-person household with PV panels, gas heating, and electric hot water. Structural changes include purchasing an electric vehicle, introducing electric space heating, the arrival of a new child, and installing an electrically heated swimming pool.
- **LUCID_2 (Fig. 2b)**: A 5-person household without PV or electric vehicle, with gas heating and gas domestic hot water. Structural changes include installing PV, converting hot water to electricity, deploying smart appliance control, purchasing an electric vehicle, adding other PV, installing electric pool heating, and reducing occupancy when children leave home.
- **LUCID_3 (Fig. 2c)**: A 4-person household with an initial electric vehicle but no PV, and gas heating and hot water. Changes include installing PV, reducing occupancy when one child leaves home, temporary discontinuation of electric-vehicle use, deployment of smart-appliance control, installation of electric central heating, and reintroduction of an electric vehicle.

[2] `https://www.iseesystems.com/`

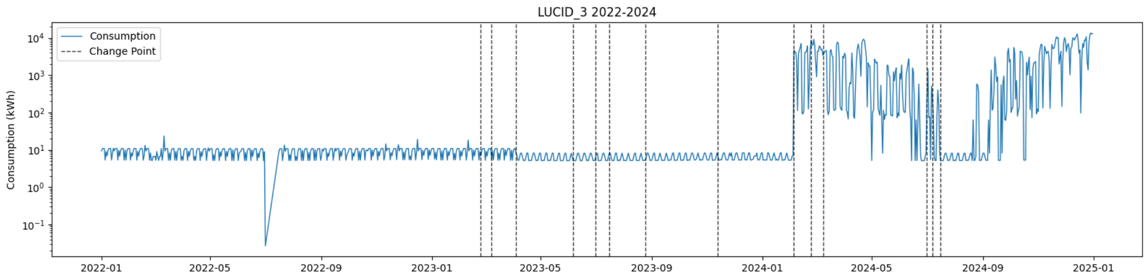
Each profile includes a ground-truth change-point indicator. Because multiple changes can accumulate, these provide a richer and more complex benchmark for CPD in realistic residential contexts. As presented in Figures 2a–2c, each household exhibits distinct consumption patterns and cumulative structural changes over the three-year observation period; profiles are visualized on a logarithmic scale for improved readability across the wide range of household electricity loads.



(a) LUCID_1 2022-2024



(b) LUCID_2 2022-2024



(c) LUCID_3 2022-2024

Figure 2: LUCID datasets

3. Results and Discussion

3.1. Forecasting Performance Overview

Table 1a reports the best forecasting model per dataset according to MAPE. Two observations are relevant for CPD. First, XGB is the most competitive model on the three ASHRAE datasets and on LUCID_2. However, the most accurate forecasting model is not necessarily the best one for forecast-driven CPD: if forecasts are too accurate, residual errors can become too small and smooth, which weakens the contrast needed to detect structural changes. Second, very large MAPE values occur on LUCID_1 for multiple models, indicating strong denominator sensitivity when true consumption approaches zero.

Table 1: Forecast and CPD Results

(a) Best forecast model per dataset

Dataset	Best model	MAPE
Apartment Mid-Rise	XGB	0.0480
Small Office	XGB	0.3749
Sit-Down Restaurant	XGB	0.0317
LUCID_1	KNN	9.81×10^{14}
LUCID_2	XGB	0.1328
LUCID_3	KNN	0.7501

(b) Comparison counts between residual-based and raw CPD.

Metric	Residual better	Raw better	Tie	Residual win rate
Covering	474	472	2330	14.47%
F1 Score	158	494	2624	4.82%
G-mean	158	494	2624	4.82%

3.2. Sensitivity to Tolerance Window

Table 1b summarizes all pairwise comparisons between raw CPD and residual CPD for the tested metrics across 3276 configurations, computed as 6 datasets $\times$ 14 forecasting models $\times$ 13 CPD algorithms $\times$ 3 tolerance windows. This global view shows a large number of ties and an overall advantage of raw CPD on pointwise detection metrics. However, the Covering metric is nearly balanced, slightly favouring forecast-driven CPD suggesting that residual-based CPD can still improve segmentation quality in specific contexts even when pointwise gains are limited.

To examine how this balance depends on temporal matching tolerance, results are disaggregated by: 1) $\delta = 1$: 235 residual wins, 479 raw wins, 2562 ties; 2) $\delta = 3$: 270 residual wins, 515 raw wins, 2491 ties; 3) $\delta = 7$: 285 residual wins, 466 raw wins, 2525 ties.

Residual-based CPD improves as matching tolerance increases, while raw CPD remains dominant in strict pointwise scoring. This behaviour is consistent with the hypothesis that residual preprocessing can improve temporal localization robustness without systematically improving exact-hit detection. Since $\delta = 7$ gives the best overall results in the experiments, only results at $\delta = 7$ are presented in the remainder of the paper.

3.3. Algorithm-Specific behaviour ($\delta = 7$)

At $\delta = 7$ days, the aggregate counts over F1 Score, G-mean, and Covering are: 285 residual wins, 466 raw wins, and 2525 ties. The strongest residual gains are concentrated in a subset of CPD algorithms:

1. PELT: 68 residual wins,
2. Window: 68 residual wins,
3. Binseg: 66 residual wins,
4. KernelCPD: 44 residual wins,
5. BottomUp: 39 residual wins.

In contrast, CUSUM, EWMA, TwoSample, cpm1B, cpm1S, cpmMS, sbs, and wbs show only ties in this aggregate view. This indicates that the benefit of forecast-driven CPD is algorithm-dependent rather than universal.

3.4. Raw-vs-Residual CPD Comparison

The detailed results are shown in Table 2 and 3. There is only one matrix for F1 Score and G-mean as results are identical for both metrics. CUSUM, EWMA, TwoSample, cpm1B, cpm1S, cpmMS, sbs, and wbs are omitted from the matrices because they yield only ties.

Each cell reports three counts in the order: ■ *forecast-driven wins* / ■ *tie* / ■ *raw wins*. Because each dataset family (EnergyPlus or LUCID) contains three datasets, each cell sums up to three valid comparisons for the selected metric at $\delta = 7$. Column headers use the abbreviations BU=BottomUp and ker=KernelCPD.

Table 2: Forecast-Driven vs. Raw CPD Matrix – Covering

(a) EnergyPlus

	Binseg	BU	Ker	PELT	Window	Total
ARIMA	0/1/2	1/0/2	0/1/2	0/0/3	1/1/1	2/3/10
KNN	2/0/1	0/0/3	2/0/1	2/0/1	2/0/1	8/0/7
KNN-MLP	0/0/3	1/0/2	1/1/1	0/1/2	1/1/1	3/3/9
Lasso	1/0/2	2/0/1	1/0/2	1/0/2	2/1/0	7/1/7
Lasso-MLP	2/0/1	1/0/2	1/0/2	2/0/1	3/0/0	9/0/6
Lasso-XGB	2/0/1	1/0/2	1/0/2	2/0/1	2/0/1	8/0/7
LR	1/0/2	2/0/1	1/0/2	1/0/2	2/1/0	7/1/7
LR-XGB	2/0/1	1/0/2	1/0/2	2/0/1	2/0/1	8/0/7
LSVR	0/3/0	0/3/0	0/3/0	0/3/0	0/3/0	0/15/0
MLP	0/2/1	0/2/1	0/2/1	0/2/1	0/2/1	0/10/5
SARIMAX	0/1/2	1/0/2	0/1/2	0/0/3	1/1/1	2/3/10
SGD	2/0/1	0/0/3	1/0/2	2/0/1	0/1/2	5/1/9
SVR	2/0/1	1/1/1	0/1/2	2/0/1	1/2/0	6/4/5
XGB	3/0/0	0/0/3	2/0/1	3/0/0	1/1/1	9/1/5
Total	17/7/18	11/6/25	11/9/22	17/6/19	18/14/10	74/42/94

(b) LUCID

	Binseg	BU	Ker	PELT	Window	Total
ARIMA	0/0/3	0/3/0	1/1/1	1/1/1	1/1/1	3/6/6
KNN	2/0/1	0/3/0	1/2/0	0/2/1	1/2/0	4/9/2
KNN-MLP	2/0/1	0/3/0	2/1/0	1/1/1	2/1/0	7/6/2
Lasso	2/0/1	1/2/0	1/0/2	2/0/1	3/0/0	9/2/4
Lasso-MLP	2/0/1	0/3/0	1/1/1	2/0/1	2/0/1	7/4/4
Lasso-XGB	1/1/1	0/3/0	1/1/1	2/1/0	2/0/1	6/6/3
LR	2/0/1	2/1/0	1/0/2	3/0/0	2/0/1	10/1/4
LR-XGB	1/0/2	1/2/0	1/1/1	2/0/1	2/0/1	7/3/5
LSVR	1/0/2	0/3/0	3/0/0	2/0/1	2/0/1	8/3/4
MLP	2/0/1	0/3/0	2/0/1	2/0/1	2/0/1	8/3/4
SARIMAX	0/0/3	0/3/0	1/1/1	1/1/1	1/1/1	3/6/6
SGD	1/0/2	1/2/0	0/0/3	2/0/1	1/1/1	5/3/7
SVR	2/0/1	1/2/0	3/0/0	1/1/1	2/0/1	9/3/3
XGB	1/1/1	0/3/0	1/1/1	0/2/1	1/1/1	3/8/4
Total	19/2/21	6/36/0	19/9/14	21/9/12	24/7/11	89/63/58

Table 3: Forecast-Driven vs. Raw CPD Matrix – F1 Score & G-mean

(a) EnergyPlus

	Binseg	BU	Ker	PELT	Window	Total
ARIMA	0/2/1	2/1/0	0/2/1	1/2/0	0/2/1	3/9/3
KNN	0/0/3	1/2/0	0/0/3	0/0/3	0/0/3	1/2/12
KNN-MLP	1/2/0	2/1/0	0/3/0	1/2/0	0/3/0	4/11/0
Lasso	0/0/3	0/2/1	0/0/3	0/0/3	0/0/3	0/2/13
Lasso-MLP	0/1/2	0/2/1	0/1/2	0/1/2	0/1/2	0/6/9
Lasso-XGB	0/1/2	0/2/1	0/1/2	0/1/2	0/0/3	0/5/10
LR	0/0/3	0/2/1	0/0/3	0/0/3	0/0/3	0/2/13
LR-XGB	0/0/3	0/2/1	0/0/3	0/0/3	0/0/3	0/2/13
LSVR	0/3/0	0/3/0	0/3/0	0/3/0	0/3/0	0/15/0
MLP	0/2/1	1/2/0	0/2/1	0/2/1	0/2/1	1/10/4
SARIMAX	0/2/1	2/1/0	0/2/1	1/2/0	0/2/1	3/9/3
SGD	0/0/3	0/2/1	0/0/3	0/0/3	0/0/3	0/2/13
SVR	1/0/2	1/2/0	1/1/1	1/1/1	0/2/1	4/6/5
XGB	0/0/3	0/2/1	0/0/3	0/0/3	0/1/2	0/3/12
Total	2/13/27	9/26/7	1/15/26	4/14/24	0/16/26	16/84/110

(b) LUCID

	Binseg	BU	Ker	PELT	Window	Total
ARIMA	1/2/0	0/3/0	1/2/0	1/2/0	1/1/1	4/10/1
KNN	0/3/0	0/3/0	0/3/0	0/3/0	0/2/1	0/14/1
KNN-MLP	1/1/1	0/3/0	0/3/0	0/2/1	2/1/0	3/10/2
Lasso	0/1/2	0/2/1	0/1/2	0/2/1	1/1/1	1/7/7
Lasso-MLP	1/1/1	0/3/0	0/2/1	0/2/1	1/0/2	2/8/5
Lasso-XGB	1/1/1	0/3/0	0/2/1	0/2/1	0/1/2	1/9/5
LR	2/0/1	0/3/0	1/0/2	2/0/1	1/0/2	6/3/6
LR-XGB	1/1/1	0/3/0	0/2/1	1/1/1	0/1/2	2/8/5
LSVR	1/1/1	1/2/0	2/1/0	2/1/0	2/1/0	8/6/1
MLP	2/1/0	0/2/1	1/0/2	2/1/0	3/0/0	8/4/3
SARIMAX	1/2/0	0/3/0	1/2/0	1/2/0	1/1/1	4/10/1
SGD	0/2/1	0/3/0	0/2/1	1/1/1	0/1/2	1/9/5
SVR	2/1/0	1/2/0	0/3/0	1/2/0	0/2/1	4/10/1
XGB	0/2/1	0/3/0	0/2/1	0/2/1	1/1/1	1/10/4
Total	13/19/10	2/38/2	6/25/11	11/23/8	13/13/16	45/118/47

3.5. Threats to Validity

Dataset realism. The study relies on synthetic datasets to ensure exact ground-truth change points. This is necessary for objective CPD evaluation, but it cannot fully reproduce all field conditions (sensor drift, missing data, undocumented interventions, occupant irregularities, and control overrides).
Tolerance-window sensitivity. CPD metrics depend on the matching window δ. The observed shift from $\delta = 1$ to $\delta = 7$ shows that conclusions vary with temporal tolerance.
Forecast-to-CPD transfer gap. Better forecasting accuracy does not necessarily improve CPD. Residuals can become too smooth or weakly contrasted, reducing CP separability .
High tie prevalence and algorithm degeneracy. The large number of ties can mask weak discrimination power. In addition, several CPD families are tie-dominant across settings, limiting practical gains from preprocessing.
Metric instability near low loads. Very large MAPE values indicate denominator sensitivity and limit the interpretation of forecast-quality rankings when consumption approaches zero.

3.6. Practical Implications

For deployment in building and district-level energy management, forecast-driven CPD should be configured as an adaptive layer rather than a mandatory pre-processing step. Algorithm selection and tolerance settings should be tuned per asset type and operational objective (early warning versus precise change localization).

4. Conclusion and Perspectives

This work investigated forecast-driven CPD for electricity consumption time series, with a focus on realistic seasonality and controlled ground-truth changes. The proposed framework separates predictable demand dynamics from structural deviations by applying CPD to residuals.

Experimental evidence shows that forecast-driven CPD is beneficial in a subset of configurations, but not universally across algorithms, datasets, and metrics. Raw CPD remains competitive, and ties are frequent. The conclusion is then conditional: residual-based preprocessing can improve robustness and segmentation quality in specific regimes, but it should not be assumed to dominate by default.

Several concrete examples support this conclusion. First, better forecasting does not automatically imply better CPD: although XGB is often the best predictor by MAPE (Fig. 1a), downstream CPD gains remain concentrated in specific forecast-model/CPD pairs rather than being systematic. Second, the conclusion is metric-dependent: strict pointwise metrics favor raw CPD, while Covering is nearly balanced globally (Fig. 1b). Third, the conclusion is CPD-algorithm dependent, with the most recurrent residual gains for PELT, Binseg, Window, BottomUp, and KernelCPD, while several other methods remain tie- dominant. Fourth, performance is tolerance-dependent: increasing δ from 1 to 7 raises residual wins from 497 to 601 and reduces raw wins from 1311 to 1250.

These patterns are consistent with the matrix visualizations for both dataset families and all retained metrics (Table 2 and 3).

Future work will address three priorities: 1) extension to additional real datasets with partial labeling, 2) probabilistic and uncertainty-aware forecasting for more reliable residual modeling, and 3) online CPD configurations for near-real-time building and district monitoring.

Acknowledgments

We thank the financial support from the Service Public de Wallonie Recherche under Grant No. 2110107 - SERENITY2 by WIN2WAL. We thank X. Calixte for the creation of the LUCID datasets in the SERENITY research project. We acknowledge the Brussels-Capital Region - Innoviris (Brussels Public Organisation for Research and Innovation) for financial support under grant number 2021-RDIR-49b. This work was also partially supported by: the AIDE project funded by the Belgian SPF BOSA under the program “Financing of projects for the development of artificial intelligence in Belgium” with reference number 06.40.32.33.00.10; and the São Paulo Research Foundation (FAPESP) – grants 2017/21174-8 and 2020/00123-9. Computational resources have been provided by the *Consortium des Équipements de Calcul Intensif* (CÉCI), funded by the *Fonds de la Recherche Scientifique de Belgique* (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

Nomenclature

y_t observed electricity consumption at time t
$\hat{y}_t$ forecasted electricity consumption at time t
ε_t residual at time t, with $\varepsilon_t = y_t - \hat{y}_t$
t time index
τ change-point index
δ tolerance window (days) for CP matching

References

[1] Aminikhanghahi S., Cook D.J. *A survey of methods for time series change point detection*. Knowledge and Information Systems 2017.

[2] Basseville M, Nikiforov IV. *Detection of abrupt changes: theory and application*. Englewood Cliffs: Prentice hall; 1993.

[3] Bottou L. *Large-scale machine learning with stochastic gradient descent*. In: Lechevallier Y., Saporta G., editors. COMPSTAT 2010: Proc. of the 19th Int. Conf. on Computational Statistics; 2010 Aug 22-27; Paris, France. Heidelberg, Germany: Physica-Verlag; 2010.

[4] Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. *Time series analysis: forecasting and control*. 5th ed. Hoboken, NJ, USA: Wiley; 2015.

[5] Celisse A, Marot G, Pierre-Jean M, Rigaill GJ. *New efficient algorithms for multiple change-point detection with reproducing kernels*. Computational Statistics & Data Analysis. 2018.

[6] Chen T., Guestrin C. *XGBoost: A scalable tree boosting system*. In: Proc. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining; 2016 Aug 13-17; San Francisco, CA, USA. New York, NY, USA: ACM; 2016.

[7] Cover T., Hart P. *Nearest neighbor pattern classification*. IEEE Trans Inf Theory 1967.

[8] Fenza G., Gallo M., Loia V. *Drift-Aware Methodology for Anomaly Detection in Smart Grid*. IEEE Access 2019.

[9] Fryzlewicz P. *Wild binary segmentation for multiple change-point detection*. Ann Stat 2014.

[10] Gaur M., Makonin S., Baji'c I.V., Majumdar A. *Performance Evaluation of Techniques for Identifying Abnormal Energy Consumption in Buildings*. IEEE Access 2019.

[11] Gulati M., Arjunan P. *LEAD1.0: A Large-scale Annotated Dataset for Energy Anomaly Detection in Commercial Buildings*. In: Proc. of the Thirteenth ACM Int. Conf. on Future Energy Systems; 2022 Jun 28–Jul 1; Virtual Event, USA. New York, NY, USA: ACM; 2022.

[12] Heo Y., Zavala V.M. *Gaussian process modeling for measurement and verification of building energy savings*. Energy and Buildings 2012.

[13] Heidrich B., Ludwig N., Turowski M., Mikut R., Hagenmeyer V. *Adaptively coping with concept drifts in energy time series forecasting using profiles*. In: Proc. of the Thirteenth ACM Int. Conf. on Future Energy Systems; 2022 Jun 28–Jul 1; Virtual Event, USA. New York, NY, USA: ACM; 2022.

[14] Hong T., Pinson P., Wang Y., Weron R., Yang D., Zareipour H. *Energy Forecasting: A Review and Outlook*. IEEE Open Access Journal of Power and Energy 2020.

[15] Hossain M.F. *Advanced Building Design*. In: Hossain M.F., editor. Sustainable Design and Build. Oxford, UK: Butterworth-Heinemann; 2019.

[16] Hyndman R.J., Koehler A.B. *Another look at measures of forecast accuracy*. Int. J. Forecasting 2006.

[17] Keogh E, Chu S, Hart D, Pazzani M. *An online algorithm for segmenting time series*. In Proc. 2001 IEEE Int. Conf. on Data Mining; 2001 Nov 29; IEEE).

[18] Killick R, Fearnhead P, Eckley IA. *Optimal detection of changepoints with a linear computational cost*. J. Am. Stat. Assoc. 2012.

[19] Killick R., Eckley I.A. *changepoint: An R Package for Changepoint Analysis*. J. Stat. Softw. 2014.

[20] Lin G., Claridge D.E. *A temperature-based approach to detect abnormal building energy consumption*. Energy and Buildings 2015.

[21] Calixte X., de Schietere de Lophem M. *LUCID Synthetic Residential Electricity Consumption Profiles 2022-2024*. Louvain-la-Neuve, Belgium: Open Data @ UCLouvain; 2026. Available at: `https://doi.org/10.14428/DVN/FRPFHK`

[22] Luo J., Hong T., Yue M. *Real-time anomaly detection for very short-term load forecasting*. Journal of Modern Power Systems and Clean Energy 2018.

[23] Nascimento G.F.M., Wurtz F., Kuo-Peng P., Delinchant B., Batistela N.J. *Outlier Detection in Buildings' Power Consumption Data Using Forecast Error*. Energies 2021.

[24] Neter J., Wasserman W., Kutner M.H. *Applied linear regression models*. Homewood, IL, USA: Irwin; 1983.

[25] Nikseresht A., Amindavar H. *Energy demand forecasting using adaptive ARFIMA based on a novel dynamic structural break detection framework*. Applied Energy 2024.

[26] Obst D., de Vilmarest J., Goude Y. *Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France*. IEEE Trans. on Power Systems 2020.

[27] Page E.S. *Continuous inspection schemes*. Biometrika 1954.

[28] Roberts S.W. *Control chart tests based on geometric moving averages*. Technometrics 1959.

[29] Ross G.J., Tasoulis D.K., Adams N.M. *Nonparametric monitoring of data streams for changes in location and scale*. Technometrics 2011.

[30] Ross G.J. *Parametric and nonparametric sequential change detection in R: The cpm package*. J. Stat. Softw. 2015.

[31] Rumelhart D.E., Hinton G.E., Williams R.J. *Learning representations by back-propagating errors*. Nature 1986.

[32] Scott AJ, Knott M. *A cluster analysis method for grouping means in the analysis of variance*. Biometrics 1974.

[33] Shumway R.H., Stoffer D.S. *Time series analysis and its applications*. New York, NY, USA: Springer; 2000.

[34] Smola A.J., Schölkopf B. *A tutorial on support vector regression*. Stat Comput 2004.

[35] Suykens J.A.K., Van Gestel T., De Brabanter J., De Moor B., Vandewalle J. *Least squares support vector machines*. Singapore: World Scientific; 2002.

[36] Takahashi K., Ooka R., Kurosaki A. *Seasonal threshold to reduce false positives for prediction-based outlier detection in building energy data*. Journal of Building Engineering 2024.

[37] Tibshirani R. *Regression shrinkage and selection via the lasso*. J Roy Stat Soc B 1996.

[38] Touzani S., Ravache B., Crowe E., Granderson J. *Statistical change detection of building energy consumption: Applications to savings estimation*. Energy and Buildings 2019.

[39] Truong C., Oudre L., Vayatis N. *A review of change point detection methods*. arXiv 2018.

[40] Truong C., Oudre L., Vayatis N. *Selective review of offline change point detection methods*. Signal Processing 2020.

[41] Truong C., Oudre L., Vayatis N. *ruptures: change point detection in Python*. arXiv 2018.

[42] Turowski M., Weber M., Neumann O., Heidrich B., Phipps K., Çakmak H.K., Mikut R., Hagenmeyer V. *Modeling and generating synthetic anomalies for energy and power time series*. In: Proc. of the Thirteenth ACM Int. Conf. on Future Energy Systems; 2022 Jun 28–Jul 1; Virtual Event, USA. New York, NY, USA: ACM; 2022.

[43] Turowski M., Heidrich B., Weing"artner L., Springer L., Phipps K., Sch"afer B., Mikut R., Hagenmeyer V. *Generating synthetic energy time series: A review*. Renewable and Sustainable Energy Reviews 2024.

[44] Van den Burg G.J.J., Williams C.K.I. *An evaluation of change point detection algorithms*. arXiv 2020.

[45] Victor Khamesi. *ocpdet: A Python package for online changepoint detection in univariate and multivariate data*. (Version v0.0.5). Zenodo 2022.