Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An IoT-Enabled Green Energy Management and Audit System for Commercial Buildings: A Simulation-Based Case Study for Uttar Pradesh, India

Tirth Raj Gautam

tirth.2024rms10@mnnit.ac.in

Motilal Nehru National Institute of Technology

TANUJ NANDAN

Motilal Nehru National Institute of Technology

NEKRAM RAWAL

Motilal Nehru National Institute of Technology

Case Report

Keywords: Green energy management, IoT, ThingSpeak, ensemble machine learning, energy audit, commercial buildings, Uttar Pradesh, difference-in-differences, simulation-based case study, Sustainable Development Goals

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11458925/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An IoT-Enabled Green Energy Management and Audit System for Commercial Buildings: A Simulation-Based Case Study for Uttar Pradesh, India

TIRTH RAJ GAUTAM[a] » Management » Research Scholar » SCHOOL OF MANAGEMENT STUDIES, MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD, PRAYAGRAJ, INDIA - 211004» tirth.2024rms10@mnnit.ac.in

Dr. TANUJ NANDAN » Management » Professor » SCHOOL OF MANAGEMENT STUDIES, MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD, PRAYAGRAJ, INDIA – 211004 » tanuj@mnnit.ac.in

Dr. NEKRAM RAWAL » Environmental Engineering » Professor » DEPARTMENT OF CIVIL ENGINEERING, MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD, PRAYAGRAJ, INDIA - 211004 » nrrawal@mnnit.ac.in

Abstract

Commercial buildings in Uttar Pradesh, India, rely largely on periodic energy audits, leaving day-to-day inefficiencies in HVAC control, lighting and equipment standby power undetected between audits. This paper presents a hypothetical, simulation-based case study of an IoT-enabled Green Energy Management and audit System (GEMS): 12 virtual commercial facilities in Lucknow, Kanpur, Varanasi and Agra were simulated at hourly resolution over 24 months, with simulated sub-metering and environmental-sensor readings carried on a conceptual ThingSpeak-style cloud layer. Under the stated waste model, 21.3% of simulated consumption was avoidable (95% CI 20.5–22.2%). A three-model ensemble (XGBoost, Random Forest, neural network) evaluated with expanding-window time-series validation detected labelled synthetic faults with 79.8% sensitivity and 76.1% precision (F1 = 0.78). A season-matched, log-space difference-in-differences comparison of six intervention and six control facilities gave 20.7% energy savings (facility-level range 15.4–28.1%; indicative bootstrap CI 17.7–24.3%, n = 6), with no material decay in the second year. Under illustrative cost and tariff assumptions, simple payback was 2.4 years, NPV at 8% was ₹130.1 million, IRR was 39.6%, and avoided emissions were about 2,964 tCO_2 per year (0.72 tCO_2/MWh). These are simulation outputs conditional on stated assumptions, not measured savings from real buildings. By quantifying avoidable commercial-building energy waste and the CO_2 emissions it implies, the study speaks directly to Sustainable Development Goal (SDG) 7 (Affordable and Clean Energy), SDG 9 (Industry, Innovation and Infrastructure), SDG 11 (Sustainable Cities and Communities) and SDG 13 (Climate Action), and is intended to give Uttar Pradesh energy planners, ESCOs and building owners an evidence-shaped starting point for pursuing these targets in the state's commercial building stock.

Keywords Green energy management · IoT · ThingSpeak · ensemble machine learning · energy audit · commercial buildings · Uttar Pradesh · difference-in-differences · simulation-based case study · Sustainable Development Goals

1. Introduction

1.1 Background and context

India's commercial building stock is expanding quickly alongside urbanisation, and buildings overall account for a large and growing share of national electricity use. Commercial-sector consumption specifically is a smaller but fast-growing slice of the total — on the order of a high single-digit percentage of all-India electricity consumption in recent years — while industry and households make up most of the remainder [1]. The Bureau of Energy

Efficiency's Energy Conservation Building Code (ECBC), first issued in 2007 and substantially revised in 2017, sets minimum energy-performance requirements for commercial buildings above a defined connected-load threshold, and a majority of Indian states, Uttar Pradesh included, have formally notified some version of the code. Adoption of the code into everyday design and operating practice nonetheless remains uneven, and enforcement capacity varies a great deal across states and municipal bodies [2].

Traditional energy audits — typically annual or biennial walk-through or Level-I/II/III exercises under ISO 50002 — give a useful snapshot but cannot see the day-to-day drift that actually drains a building's energy budget: a chiller setpoint left too low overnight, lights running in an empty conference room, a server room's standby draw creeping up after a firmware update [3]. Continuous, sensor-based monitoring closes exactly this temporal gap. Internet of Things (IoT) architectures — distributed sensors, a lightweight edge controller, a cloud data layer and an analytics/alerting front end — are now a well-established way to build that continuous view in both the smart-city and building-management literatures [4].

1.2 Motivation and problem statement

Building energy management systems (BMS) that combine sub-metering with automated control are effective where they exist, but comprehensive BMS retrofits are capital-intensive and are mostly found in premium commercial developments; the wider stock of mid-size Indian commercial buildings — the offices, retail units, nursing homes, hotels and schools that make up most of Uttar Pradesh's commercial floor area — is largely unmetered beyond the utility meter at the point of supply. A lighter-weight monitoring-and-audit layer, built on commodity IoT hardware and a hosted IoT cloud platform rather than a full building-automation retrofit, is a more plausible near-term step for this segment, but there is limited published evidence — for India specifically — of what such a system could plausibly achieve, at what cost, and with what degree of statistical confidence.

This paper addresses that gap the same way the companion smart-waste-collection paper in this research programme [15] did: not with a physical pilot, but with a controlled, fully-specified simulation that lets every assumption be stated explicitly and every result be reproduced from a documented pipeline. The purpose is to give municipal energy planners, ESCOs and building owners in Uttar Pradesh an evidence-shaped — but honestly labelled as simulated — starting point for their own techno-economic thinking, not a claim of measured field performance. Framed this way, the exercise is also a direct, India-specific contribution to the Sustainable Development Goals: commercial-building energy waste is a tangible barrier to SDG 7 (ensuring access to affordable, reliable, sustainable and modern energy), and closing it through low-cost continuous monitoring rather than capital-intensive retrofit speaks to SDG 9's call for resilient, resource-efficient infrastructure and SDG 11's target of making cities and human settlements more sustainable; the avoided-emissions result in Section 4(e) is this paper's quantified link to SDG 13 (climate action).

1.3 Objectives of this simulation study

The study was designed around four questions, each answered from the simulation pipeline described in Section 3 rather than assumed in advance:

- RQ1 — Waste characterisation: under a stated end-use split (HVAC, lighting, equipment standby, thermal envelope) adopted as a design assumption informed by the literature, what share of simulated commercial-building energy use is avoidable, and how does it vary by building type and by the Uttar Pradesh seasonal cycle?
- RQ2 — Detection performance: can an ensemble of standard, reproducible machine-learning models (gradient boosting, random forest, a compact neural network), evaluated with time-series-correct cross-validation, detect injected, labelled consumption anomalies at a useful sensitivity and precision?
- RQ3 — Realised savings and persistence: in a simulated six-vs-six control/intervention comparison, what energy-saving effect emerges once a season-matched difference-in-differences correction is applied, and does the effect persist or decay across the second simulated year?

- RQ4 — Techno-economic viability: under stated Indian cost, tariff and discount-rate assumptions, what capital cost, payback period, NPV and CO_2 reduction does the simulated intervention imply, and how sensitive is that conclusion to the underlying assumptions?

1.4 Relevance to the Sustainable Development Goals

Each of the four research questions above is also a question about sustainable development. RQ1 (waste characterisation) sizes the SDG 7 and SDG 13 opportunity available to Uttar Pradesh's commercial building stock by quantifying how much energy consumption is avoidable before any capital is spent. RQ2 (detection performance) asks whether that opportunity can be found and acted on continuously, which is the operational precondition for SDG 9's resource-efficiency agenda rather than a one-off audit exercise. RQ3 (realised savings and persistence) tests whether a lighter-weight monitoring layer can deliver savings that persist, which is what distinguishes a genuine contribution to SDG 11's sustainable-cities target from a short-lived pilot effect. RQ4 (techno-economic viability) asks whether the intervention is financially self-sustaining under Indian tariffs and costs, because a decarbonisation pathway that is not bankable for building owners and ESCOs is not a pathway state policy can scale. Section 4(e) reports the resulting avoided-emissions estimate, and Section 7.2 returns to the policy implications of these SDG linkages.

2. Literature Review

2.1 Energy use and efficiency potential in commercial buildings

Pérez-Lombard, Ortiz and Pout's widely cited review of building energy-consumption information sets out the now-standard end-use decomposition used across the field — HVAC, lighting, plug/equipment loads and other systems — and is the basis for the illustrative end-use split adopted in Section 3.3.1 of this paper, as a design assumption rather than a measured value (the review draws mainly on US and European building stock) [5]. Systematic commissioning and retro-commissioning studies in developed-economy building stocks, surveyed by Mills, report median whole-building savings in roughly the 5–20% range at attractive paybacks, with existing (as opposed to new-construction) buildings tending toward the higher end because more low-cost operational waste has typically accumulated [6]. Granderson and Lin's synthesis of building energy-information-system deployments finds that monitoring-and-analytics layers of the kind simulated here typically unlock savings in a broadly similar band once alerting and operator follow-through are in place, and stress that realised savings depend heavily on organisational follow-through, not the technology alone [7] — a dependency this paper's simulation design deliberately encodes through a facility-level “engagement” parameter (Section 3.7).

2.2 IoT architectures and cloud platforms for monitoring

Zanella, Bui, Castellani, Vangelista and Zorzi's IEEE Internet of Things Journal paper on IoT for smart cities remains a standard reference for the sensor → gateway → cloud → application architecture pattern used across smart-building and smart-city monitoring systems, including the one simulated in this paper [4]. ThingSpeak, the specific hosted IoT platform used as the conceptual cloud layer here (and in the companion waste-collection paper this study follows), is a MathWorks-operated service offering channel/field-based time-series storage, a REST API, and native MATLAB-based analytics; its free-tier service enforces an update interval of at least 15 seconds between writes to a channel, a cap of 3 million messages per year and a limit of four channels, and each channel supports up to eight numeric data fields plus location and status metadata [8]. The update-interval and message-volume limits are well above what an hourly-aggregated commercial-building monitoring layer of the kind simulated here would require (twelve channels written hourly generate roughly 0.1 million messages per year), and the eight-field-per-channel ceiling is respected in the channel design set out in Section 3.3. The four-channel cap would, however, be exceeded by the twelve-channel design simulated here, so a real deployment would need an academic or standard licence.

2.3 Machine learning for load forecasting and anomaly detection

Gradient-boosted trees and random forests are established, well-benchmarked baselines for short-horizon building-load forecasting; Chen and Guestrin's XGBoost paper and Breiman's original random-forest paper underpin the two tree-based ensemble members used in Section 3.5 [9], [10]. Forecast-residual methods — flagging a reading as anomalous when it deviates from a model's expectation by more than a noise-scaled threshold — are a standard, general anomaly-detection approach and are the basis for the detection logic used here; the companion literature on time-series cross-validation cautions that a naive random train/test split on time-ordered data allows future information to leak into training and inflates reported accuracy, which is why an expanding-window scheme is used throughout Section 4.2 instead [11].

2.4 Energy-management standards and policy context in India

ISO 50001 (energy management systems) and the associated ISO 50002 energy-audit standard define the formal, three-tier (preliminary / detailed / investment-grade) audit framework referenced throughout this paper's discussion of what continuous monitoring adds beyond a periodic audit [3]. In India, the Bureau of Energy Efficiency's Energy Conservation Building Code sets the applicable minimum-performance benchmark for larger commercial buildings [2], and the Ministry of Power's standard default-setting advisory for room air-conditioners — 24 °C, with an approximate 6% energy saving typically cited for each degree the setpoint is raised toward that default — is cited here only as an order-of-magnitude reference point [12]; the HVAC control-loss coefficient used in the simulation's waste model (Section 3.3.1, Eq. 3) is a separate design assumption set well above that per-degree figure.

Across this literature, three gaps motivate the present study specifically: (i) published, India-specific quantification of commercial building energy waste and its end-use composition is limited; (ii) building-energy machine-learning studies do not consistently use leakage-safe time-series validation; and (iii) very few studies make explicit, and test the sensitivity of, the economic assumptions (tariff, discount rate, capital cost per unit area) that ultimately determine whether a project is viable. This paper's simulation design (Section 3) is built specifically to address all three within an honestly-labelled hypothetical case study rather than to claim field-measured results.

2.5 Building energy management and the Sustainable Development Goals

The building-efficiency and IoT-monitoring literature reviewed above (Sections 2.1–2.3) is increasingly framed explicitly around the SDGs rather than treated as a purely technical or cost-saving exercise. Commercial-building energy waste is a direct obstacle to SDG 7's target of universal access to affordable, reliable and modern energy, because wasted consumption raises the effective cost of energy for everyone else drawing on the same grid; the commissioning and energy-information-system literature surveyed in Section 2.1 is, in effect, evidence on how much of SDG 7's efficiency component is recoverable without new generation capacity. The IoT architectures reviewed in Section 2.2 are a concrete instance of the resilient, resource-efficient infrastructure called for under SDG 9, since a lightweight sensing-and-analytics layer extends the useful life and efficiency of existing building stock rather than requiring new construction. Because India's commercial floor area is concentrated in urban Uttar Pradesh, reducing its energy intensity also speaks to SDG 11's target of sustainable cities, and the CO_2-reduction evidence this paper reports in Section 4(e) is its contribution to SDG 13 on climate action. This paper's contribution relative to this literature is to make the SDG 7/9/11/13 linkage quantitative and India-specific — through the waste, detection, savings and emissions estimates in Section 4 — rather than asserting the connection qualitatively, as much of the general building-efficiency literature does.

3. Methodology

This is a simulation-based case study, following the same construction method as the companion smart-waste-collection paper in this research programme [15]: a controlled, fully-documented simulation stands in for physical sensor deployment, and every downstream number is the actual output of running real statistical and machine-learning code against that simulated data, not an asserted figure. Table 1 summarises the overall workflow; the

rest of Section 3 documents each stage and every assumption in enough detail that the pipeline could be re-run or audited.

Table 1 Simulation workflow

Stage	What was done
1. Facility & scenario design	Twelve virtual commercial facilities specified across four UP cities and six building types; split into a 6-facility control arm and a 6-facility intervention arm (Section 3.7).
2. Synthetic sensor-data generation	Hourly, 24-month (Apr 2024–Mar 2026) consumption, temperature, occupancy and control-loop-deviation series generated per facility from stated physical and behavioural assumptions (Section 3.2).
3. Simulated cloud layer	Readings carried on a conceptual ThingSpeak-style channel/field structure with realistic update-interval and field-count constraints (Section 3.3).
4. Waste quantification & alerting	End-use waste computed hour-by-hour from the stated formulas; threshold-based alert tiers defined (Section 3.3.1, 3.4).
5. Machine-learning detection	XGBoost + Random Forest + neural-network ensemble trained with expanding-window time-series cross-validation; anomalies flagged from validation-tuned residual thresholds (Section 3.5).
6. Statistical evaluation	Season-matched, log-space difference-in-differences comparison of the control and intervention arms; bootstrap confidence intervals; stability/heterogeneity checks (Section 3.6).
7. Economic evaluation	CAPEX/OPEX cost model, NPV/IRR/payback under stated tariff and discount-rate assumptions, and a ±15% sensitivity sweep (Section 3.6).

Source: simulation design for this study, following the workflow-table convention of the companion ThingSpeak paper.

3.1 System architecture and simulation setup

No physical sensors, controllers or buildings were used. As in the companion paper's waste-collection system, the behaviour of each component — sensor sampling and drift, edge aggregation, wireless/cloud transmission, threshold alerting and analytics — was modelled directly in code, and a conceptual architecture was defined to represent a realistic IoT-based commercial energy-monitoring system rather than a specific vendor deployment. The simulated architecture comprises:

- Smart sub-metering and environmental sensors (simulated): per-facility hourly active-power, indoor/outdoor temperature-deviation, and occupancy-proxy readings, at one sensing point per roughly 160 m² of floor area.
- A virtualised edge controller (Node-MCU/ESP32-class, simulated): represents local aggregation of raw readings to hourly values before upload, consistent with the update-interval constraint in Section 3.3.
- A WiFi/cloud communication layer (simulated) carrying aggregated readings to the cloud data layer.
- A ThingSpeak-style cloud data layer (simulated) for channel storage, time-stamped records and REST-style retrieval, as described in Section 3.3.
- A web dashboard and threshold-based alert layer (Section 3.4; Fig. 9 shows a conceptual rendering).
- A machine-learning analytics layer and an economic-evaluation layer, both implemented as real, executable code against the simulated data (Sections 3.5–3.6).

3.2 Simulation of IoT sensor data

Hourly synthetic readings were generated for all twelve facilities over a 730-day (24-month) window from 1 April 2024; the analysed dataset comprises 209,376 facility-hours (727 retained days × 12 facilities × 24 hours, i.e. three of the nominal 730 days are not retained in the analysed data). Consumption for each facility-hour was built from

four physically-motivated components — a temperature-driven cooling/HVAC load (zero below a 24 °C outdoor threshold, scaled by a simulated indoor control-loop deviation), an occupancy-scaled plug-and-lighting load, a small fixed standby floor, and a modest monsoon dehumidification addition — calibrated per facility to a target annual energy-use intensity (EUI) drawn from a plausible seed range for its building type (Table 2). Key simulated parameters:

- Update interval: hourly (aggregated from an assumed higher-frequency raw sampling rate at the simulated edge controller, consistent with the ThingSpeak-style constraints in Section 3.3).
- Measurement/sensor drift noise: ±3% multiplicative noise applied to every hourly reading.
- Outdoor temperature: city-level synthetic series with a seasonal mean (18–38 °C across the Uttar Pradesh winter→summer cycle), a diurnal sinusoid and Gaussian noise, generated independently for Lucknow, Kanpur, Varanasi and Agra.
- Occupancy: a business-hours bell curve for office/retail/educational facility types, and a flatter, higher-floor curve for the round-the-clock hospital and hotel facility types, each with its own random noise draw.
- Control-loop deviation (ΔT): a simulated indoor-setpoint tracking error, feeding the HVAC waste model in Section 3.3.1.
- Labelled synthetic anomalies: four fault types — HVAC failure, lighting-control fault, equipment malfunction and sensor/data fault — injected at randomly chosen times and durations (1–13 hours) with type-specific magnitude ranges, with ground-truth labels retained for the machine-learning evaluation in Section 3.5.

3.3 Simulated ThingSpeak-style cloud integration

As in the companion paper [15], a ThingSpeak-style cloud layer is used as the conceptual IoT platform; no data are written to the real ThingSpeak service. Each of the twelve simulated facilities is modelled as carrying its own ThingSpeak-style channel, respecting the real platform's eight-numeric-field-per-channel ceiling by mapping: Field 1 active power (kW), Field 2 cumulative energy (kWh), Field 3 indoor–setpoint deviation (°C), Field 4 occupancy proxy (%), Field 5 outdoor temperature (°C), Field 6 rolling 24-hour anomaly z-score, Field 7 waste-alert tier, Field 8 device/link status. Hourly aggregation at the simulated edge controller keeps the write rate for any one channel far inside the free-tier minimum update interval of 15 seconds between writes and the associated annual message ceiling [8], with substantial headroom for higher-frequency sub-metering fields should a real deployment want them; the one free-tier limit that a twelve-channel deployment would exceed is the four-channel cap (Section 2.2). A REST-style read path (as used for the dashboard mock-up in Fig. 9) would fetch the latest N readings per field for display, exactly as the companion paper's dashboard does for bin-fill alerts.

3.3.1 Simulated waste quantification

Simulated total consumption in every facility-hour is allocated across four end uses in fixed proportions adopted as a design assumption (HVAC 44%, lighting 23%, plug/equipment 20%, other 13% of instantaneous load) following the standard end-use decomposition discussed in Section 2.1. These shares are not measured values for Uttar Pradesh buildings, and India-specific splits may differ (for example, a larger HVAC share in the hot Indo-Gangetic plain, or a smaller lighting share where LED penetration is high), so they should be replaced with BEE, CEA or building-stock survey data in any field application. A waste fraction within each end use is computed hour-by-hour from the simulated control-loop deviation and occupancy signals rather than assumed as a flat constant:

Equation (1) — End-use allocation

$$E_e(t) = s_e\, E_{\text{total}}(t), \qquad \sum_e s_e = 1$$

Equation (2) — Clip function

$$\text{clip}(x, a, b) = \min(\max(x, a),\ b)$$

Equation (3) — HVAC waste fraction

$$f_{\text{HVAC}}(t) = \text{clip}(0.17\,\Delta T(t),\ 0.08,\ 0.60)$$

Equation (4) — Lighting waste fraction

$$f_{\text{Light}}(t) = \text{clip}(0.15 + 0.55\,g_{\text{occ}}(t),\ 0.15,\ 0.68)$$

Equation (5) — Equipment-standby waste fraction

$$f_{\text{Standby}}(t) = \text{clip}\big(0.20 + 0.20\,(1 - o(t)),\ 0.20,\ 0.42\big)$$

Equation (6) — Thermal-envelope waste fraction

$$f_{\text{Env}}(t) = \text{clip}\left(0.08 + 0.06\,\frac{\max(|T(t) - 24| - 3,\ 0)}{10},\ 0.08,\ 0.34\right)$$

Equation (7) — Avoidable waste by end use

$$W_e(t) = E_e(t)\,f_e(t)$$

Equation (8) — Total simulated avoidable waste

$$W_{\text{total}} = \frac{\sum_t \sum_e W_e\,(t)}{\sum_t \sum_e E_e\,(t)} \times 100\%$$

where ΔT(t) is the simulated control-loop (setpoint-tracking) deviation in °C; o(t) ∈ [0, 1] is the occupancy proxy; $g_{occ}(t)$ is the occupancy gap (a space that is lit but under-occupied); T(t) is the outdoor temperature in °C; s_e is the end-use share; $E_e(t)$ is the end-use energy; and $f_e(t)$ is the waste fraction of end use e, given by Equations (3)–(6) for HVAC, lighting, equipment standby and thermal envelope respectively.

- HVAC waste fraction = clip(0.17×ΔT, 0.08, 0.60) — poor setpoint tracking inflates avoidable HVAC draw. The 0.17-per-°C coefficient is a simulation-design assumption; it is not derived from the BEE 24 °C default-setting advisory, whose approximate 6%-per-degree saving figure [12] describes a deliberate setpoint change and is roughly one-third of the coefficient used here. ΔT in this model is a control-loop tracking error rather than a setpoint change, and the coefficient has not been calibrated against measured data. It is therefore a high-side assumption, and the HVAC share of waste (41.5%) and the quantities that depend on it should be read accordingly (Section 6).
- Lighting waste fraction = clip(0.15 + 0.55×(occupancy gap), 0.15, 0.68) — higher when a space is simulated as lit but under-occupied, reflecting the absence of automatic occupancy-linked control assumed for the baseline (pre-intervention) condition.
- Equipment-standby waste fraction = clip(0.20 + 0.20×(1−occupancy), 0.20, 0.42).
- Thermal-envelope waste fraction = clip(0.08 + 0.06×max(|T−24|−3,0)/10, 0.08, 0.34).

These coefficients, like the end-use shares above, are simulation-design assumptions, informed by but not calibrated to the literature ranges in Section 2.1, and are stated here in full so the resulting waste percentages in Section 4(a) can be traced back to them.

3.4 Alert generation logic

Two independent threshold layers were simulated, mirroring the companion paper's alert design: a waste-level tier (based on the hour's combined waste fraction) for dashboard colour-coding, and a statistical anomaly flag (Section 3.5) for fault detection.

Equation (9) — Alert tier logic

$$A(t) = \begin{cases} \text{Green} & f_{\text{waste}}(t) < 0.15 \\ \text{Yellow} & 0.15 \le f_{\text{waste}}(t) < 0.25 \\ \text{Orange} & 0.25 \le f_{\text{waste}}(t) < 0.35 \\ \text{Red} & f_{\text{waste}}(t) \ge 0.35 \end{cases}$$

- Green (waste fraction < 15%): normal operation.
- Yellow (15–25%): monitor.
- Orange (25–35%): schedule an operational review.
- Red (> 35%): investigate immediately.

3.5 Machine-learning framework for anomaly detection

An ensemble of three standard, reproducible regressors — XGBoost, a Random Forest, and a compact feed-forward neural network (scikit-learn's MLPRegressor, two hidden layers of 48 and 24 units) — was trained to predict each facility-hour's expected consumption from calendar, weather, occupancy and lagged-consumption features (hour-of-day, day-of-week, month, weekend flag, facility code, floor area, outdoor temperature, occupancy proxy, same-hour-yesterday lag, same-hour-last-week lag, and 6-hour/24-hour trailing rolling means). No one-hour-prior lag is used, because a nearer lag would let the model partly "predict" the very anomaly it is meant to flag. The same-hour-last-week lag serves as the contamination-safe long-range history feature: a short injected anomaly (1–13 hours) cannot contaminate a value a full week back. In terms of importance, the same-hour-yesterday lag is the largest single contributor by gain and the weekly lag the second (Table 8).

Equation (10) — Anomaly residual z-score

$$z(t) = \frac{y(t) - \hat{y}(t)}{\sigma_{\text{trail}}(t)}$$

Equation (11) — Anomaly flag

$$I_{\text{anom}}(t) = \begin{cases} 1 & z(t) \geq z_{\text{thr}} \\ 0 & z(t) < z_{\text{thr}} \end{cases}$$

Model evaluation used expanding-window time-series cross-validation (train on an initial period, validate on the next chunk, then extend the training window forward) rather than a random split, so that no future information leaks into training. Ensemble weights were selected by grid search on a held-out validation slice carved from the end of the training period (never from the test period), and the anomaly-flagging threshold — a residual z-score computed against a facility's own trailing two-week noise level, itself computed from a window shifted 48 hours back so a single sustained anomaly cannot inflate its own normalising denominator — was likewise selected on that validation slice by maximising F1, then applied once, unchanged, to the held-out test period. All reported test-set metrics in Section 4(b) are therefore genuinely out-of-sample.

3.6 Economic and statistical analysis framework

The simulated intervention effect was estimated with a difference-in-differences (DiD) design: a season-matched pre-intervention baseline (the first simulated quarter, April–June 2024) is compared against later windows for both the control and intervention arms, and the control arm's own change over the identical window is subtracted from the intervention arm's change, isolating the treatment effect from the shared Uttar Pradesh seasonal cycle. Because both the seasonal swing and the treatment act multiplicatively on kWh, this subtraction is carried out in log space (so that percentage effects add rather than interact non-linearly) and converted back to a percentage for reporting; Section 4(c) documents why this matters. Confidence intervals on the headline savings figure and on the waste percentage in Section 4(a) are non-parametric bootstrap intervals (10,000 and 5,000 resamples respectively); the interval on the savings figure resamples the six intervention facilities as if they were exchangeable and is therefore indicative only (Section 4(c)). The pre-intervention parallel-trends assumption underlying the DiD design is checked directly, as a descriptive sanity check (Section 4(c)).

Equation (12) — Log-space difference-in-differences

$$D_{\log} = \left[\ln \bar{E}_{I,\text{post}} - \ln \bar{E}_{I,\text{pre}}\right] - \left[\ln \bar{E}_{C,\text{post}} - \ln \bar{E}_{C,\text{pre}}\right]$$

Equation (13) — DiD savings conversion

$$S_{\text{DiD}}\,(\%) = \left(1 - e^{D_{\log}}\right) \times 100$$

Economic evaluation used a standard discounted-cash-flow model: simulated capital expenditure (CAPEX) is a function of sensor count and floor area for each intervention facility (Table 11), annual operating expenditure (OPEX) covers cloud subscription, sensor calibration, software and a part-time energy technician, and annual savings are the simulated kWh reduction (Section 4(c)) valued at an illustrative Uttar Pradesh commercial energy tariff of ₹8.60/kWh with a ₹350/kW demand-charge assumption, broadly consistent with published UPERC commercial (LMV-2/HV-2 class) tariff orders though not tied to any single order or facility's actual bill [13]. Net present value, internal rate of return and simple/discounted payback were computed over a 10-year horizon at 6/8/10/12% discount rates, with an explicit ±15% sensitivity sweep on energy savings, tariff, CAPEX and OPEX, and a break-even-tariff search. CO_2-equivalent savings use an illustrative India grid-average emission factor of 0.72 tCO_2/MWh, of the order of magnitude reported in the Central Electricity Authority's CO_2 Baseline Database for the all-India grid in recent years [14]; the exact factor changes with each database revision and is not tied to a specific version here.

3.7 Hypothetical case study design

Twelve virtual facilities were specified across four Uttar Pradesh cities and six commercial building types, and split into a 6-facility control arm and a 6-facility intervention arm (Table 2). This is a simulated comparison, not a randomised controlled trial with real consent, recruitment or field monitoring — there is accordingly no control-group contamination survey, spot-check or consent-form procedure to report, unlike a genuine field RCT; those procedures are simply not meaningful for twelve virtual buildings and are omitted rather than fabricated. What is simulated is a facility-level engagement score (0–1) for each intervention facility, representing how actively that facility's staff are assumed to act on the system's alerts; this score drives the magnitude of the simulated savings for that facility (Section 4(c)) and lets the study examine, within the simulation, the same organisational-engagement question the wider literature raises (Section 2.1).

(Table 2, the facility roster, appears at the start of Section 4 alongside the baseline results it produced.)

4. Results and Discussion

All figures, tables and percentages below are read directly off the simulation pipeline described in Section 3 — nothing here has been separately typed in or adjusted for effect. The discussion is organised into six parts, following the companion paper's structure: (a) baseline waste characterisation, (b) machine-learning detection performance, (c) simulated intervention results, (d) economic and cost-savings analysis, (e) CO_2 reduction, and (f) a conceptual dashboard visualisation.

(a) Baseline simulated energy consumption and waste

Table 2 lists the twelve simulated facilities. Across the full 209,376-facility-hour dataset, simulated annual energy-use intensity (EUI) averaged 203.1 kWh/m²/yr (sample SD 35.1, range 144–264), consistent with the seed ranges assumed per building type. Total simulated avoidable waste (Eq. 8) was 21.3% of consumption (95% bootstrap CI 20.5–22.2%), led by HVAC control inefficiency (41.5% of total waste), lighting left on in under-occupied space (27.1%), equipment standby draw (24.8%) and thermal-envelope losses (6.6%) — see Fig. 2. This is of the same order as the 5–20% commissioning-savings range surveyed in Section 2.1, and reflects the specific waste-fraction assumptions set out in Section 3.3.1 rather than a measurement of any real building.

Table 2 Simulated facility roster and baseline results

ID	Type	City	Arm	Area (m²)	EUI (kWh/m²/yr)	Waste (%)
F1	Office	Lucknow	Control	15,000	195.8	20.5
F4	Educational	Kanpur	Control	18,000	158.8	20.5
F6	Office	Varanasi	Control	11,000	189.1	20.5
F7	Retail	Agra	Control	26,000	229.2	20.5

ID	Type	City	Arm	Area (m²)	EUI (kWh/m²/yr)	Waste (%)
F10	Retail	Kanpur	Control	20,000	223.3	20.4
F12	Hospital	Agra	Control	8,500	264.0	20.4
F2	Retail	Lucknow	Intervention	22,000	191.3	22.5
F3	Hospital	Kanpur	Intervention	12,500	229.3	19.9
F5	Hotel	Varanasi	Intervention	9,000	240.6	19.9
F8	Office	Agra	Intervention	10,500	176.1	22.8
F9	IT Park	Lucknow	Intervention	30,000	195.8	24.6
F11	Educational	Varanasi	Intervention	14,000	143.5	23.0

Source: simulation output (this study). Facilities were assigned to arms before any treatment effect was applied (Section 3.7): each of the six building types appears once in the intervention arm, whereas the control arm is not a one-to-one type match (two offices, two retail facilities, one educational facility and one hospital; no hotel or IT park).

Fig. 1 Simulated monthly portfolio energy consumption, control vs. intervention

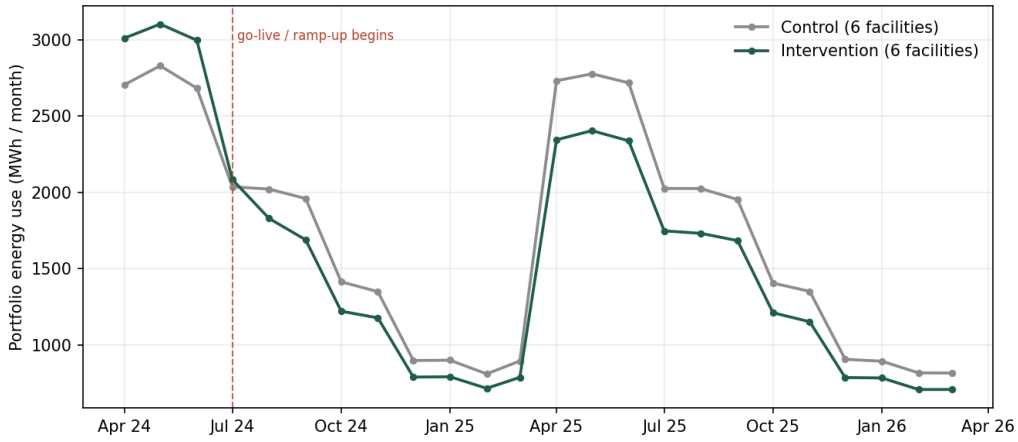


Fig. 1 Simulated monthly portfolio energy consumption, control vs. intervention arms. The shared UP seasonal cycle (summer peak, winter trough) is visible in both arms; the arms track closely before the simulated go-live (dashed line) and diverge afterward.

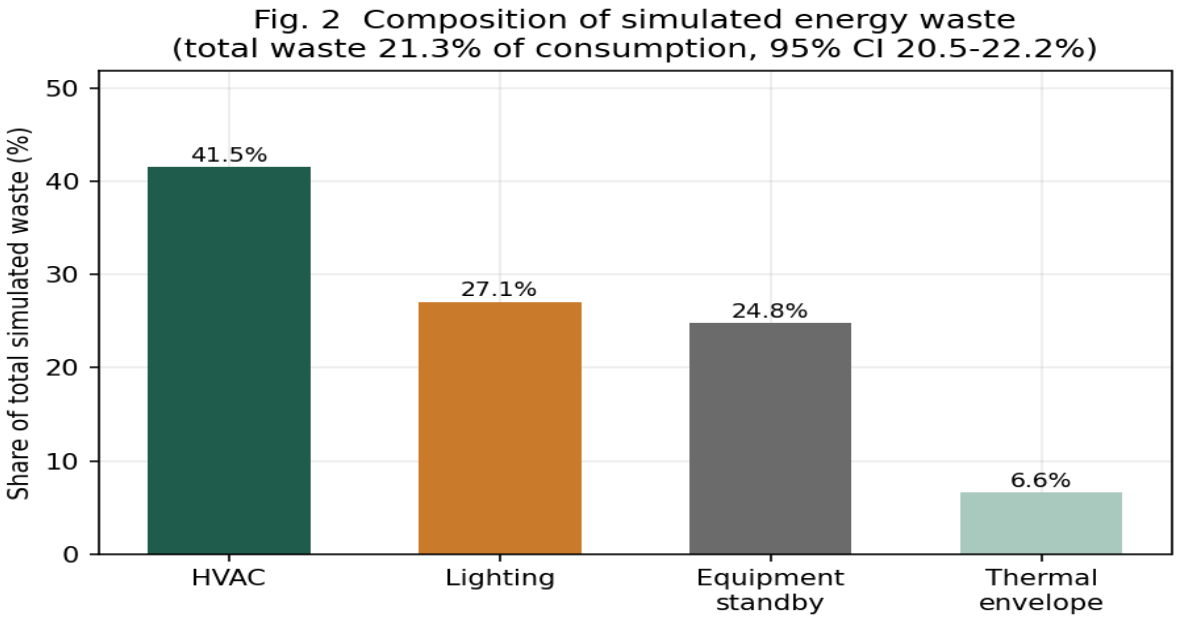


Fig. 2 Composition of simulated energy waste by end use (total 21.3% of consumption).

Table 3 Simulated waste by Uttar Pradesh season

Season	Months	Waste share of consumption (%)
Summer	Apr–Jun	22.8
Monsoon	Jul–Sep	20.7
Post-Monsoon	Oct–Nov	20.0
Winter	Dec–Mar	20.7

The simulated waste share is highest in summer, when the HVAC control-loop deviation term (Eq. 3) is both more frequent and larger in magnitude, and lowest in the post-monsoon shoulder season. Because HVAC waste depends on ΔT rather than on raw cooling load, the swing in the waste percentage across seasons is much smaller than the roughly three-fold swing in raw consumption itself (Fig. 1): a large HVAC system running efficiently in peak summer still generates proportionally less waste than a poorly-controlled one running lightly in a shoulder season.

(b) Machine-learning-based anomaly detection

The three-model ensemble (weights: XGBoost 0.5, Random Forest 0.1, neural network 0.4, selected on the validation slice described in Section 3.5) reached a held-out test-set R^2 of 0.960, RMSE 23.5 kWh/hour and MAPE 5.7%, marginally ahead of its best single member (XGBoost alone: RMSE 24.0 kWh/hour, R^2 0.958) on RMSE and R^2, but not on MAPE, where XGBoost alone is slightly better (5.5% vs. 5.7%) — a pattern suggesting that the ensemble mainly trims large absolute errors rather than percentage errors at low-load hours. It is well ahead of the naive persistence and linear-regression baselines (Table 4, Fig. 3); the ensembling gain is realistically small rather than inflated. Random Forest receives only a 0.1 weight and is the weakest single model on RMSE (28.3 kWh/hour), so it is retained for diversity rather than accuracy; the ensemble is effectively an XGBoost–neural-network blend, and a two-model ablation is a natural follow-up.

Table 4 Forecasting-model comparison, held-out test window

Model	RMSE (kWh/h)	MAPE (%)	R^2
Persistence (same hour, t-24h)	48.0	10.9	0.833
Linear Regression	49.1	23.9	0.825

Model	RMSE (kWh/h)	MAPE (%)	R^2
XGBoost (single)	24.0	5.5	0.958
Random Forest (single)	28.3	5.7	0.942
Neural Net / MLP (single)	25.9	8.2	0.951
Optimised Ensemble	23.5	5.7	0.960

Source: simulation output (this study). Ensemble weights and the anomaly threshold in Table 6 were tuned on a validation slice only; this table is the never-touched test set.

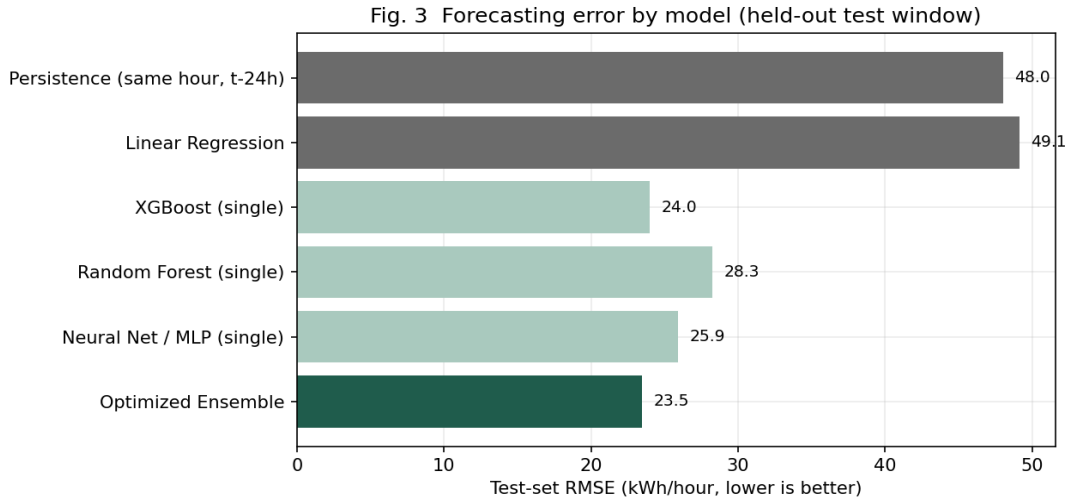


Fig. 3 Test-set forecasting error by model.

Table 5 Expanding-window time-series cross-validation

Fold	Validation rows	RMSE (kWh/h)	MAPE (%)	R^2
Fold 1	33,120	38.4	14.1	0.904
Fold 2	33,408	83.4	10.0	0.949
Fold 3	33,120	47.2	7.6	0.969
Fold 4	37,152	25.1	5.7	0.954

Each fold trains on all data up to a cut-off and validates on the following slice, so no fold ever sees its own future. Fold-to-fold RMSE varies by roughly 3× (Table 5) — fold 2, which spans part of the peak-summer ramp, is visibly harder to forecast than the calmer fold 4 — which is the expected signature of an honest time-series evaluation on seasonal data, not a modelling defect.

Anomaly detection — flagging an hour whose residual (actual − ensemble-predicted consumption) exceeds a validation-tuned, noise-normalised threshold — reached 79.8% sensitivity and 76.1% precision (F1 = 0.78, specificity 99.6%) against the 520 labelled synthetic anomalies in the test window (Table 6). Detection performance varies sensibly by fault type (Table 7, Fig. 4): the large, abrupt equipment-malfunction spikes and the near-zero sensor/data-fault dropouts are easiest to catch, sustained HVAC failures are caught at a good rate, and the comparatively small-magnitude, off-hours lighting-control faults are hardest — exactly the pattern a residual-threshold detector should show, which is itself a useful sanity check on the simulation rather than a result to be smoothed away.

Table 6 Anomaly-detection confusion matrix (test window)

	Predicted anomaly	**Predicted normal**
Actual anomaly	TP = 415	FN = 105
Actual normal	FP = 130	TN = 36,502

Threshold $z \geq 3.7$ (selected on the validation slice, applied unchanged to test).

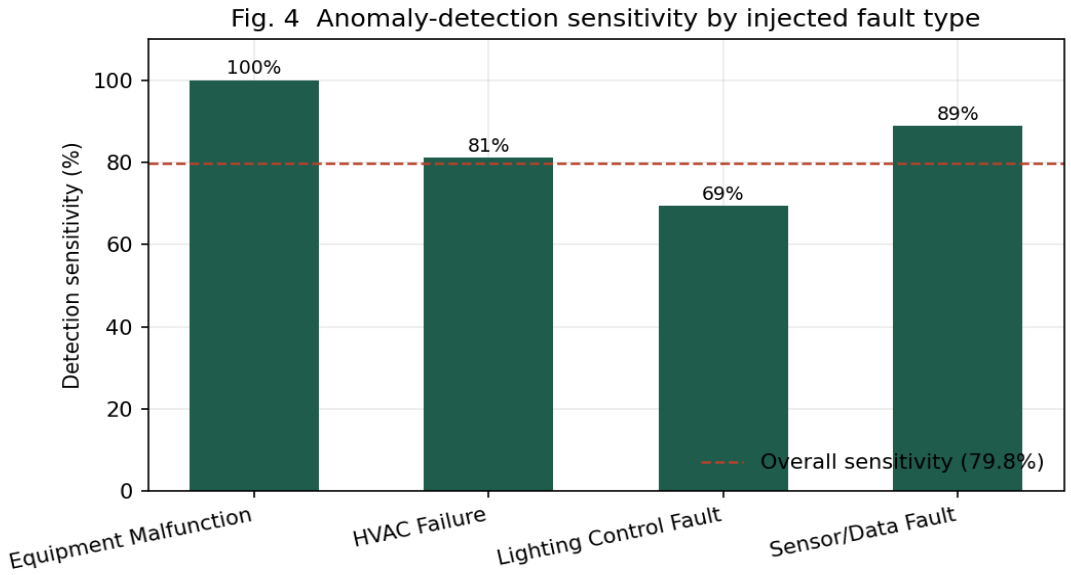


Fig. 4 Detection sensitivity by injected fault type.

Table 7 Detection by fault type

Fault type	**Injected events**	**Detected**	**Sensitivity (%)**
Equipment Malfunction	50	50	100.0
HVAC Failure	196	159	81.1
Lighting Control Fault	193	134	69.4
Sensor/Data Fault	81	72	88.9

Table 8 Feature importance (XGBoost gain, top 8)

Feature	**Importance (%)**	**Interpretation**
Same hour, 1 day prior	50.9	Largest single contributor: daily continuity
Same hour, 1 week prior	37.8	Contamination-safe long-range history: weekly operating pattern
Outdoor temperature	1.8	HVAC cooling-load driver
Floor area	1.6	Scale of facility
6-hour rolling mean	1.5	Short-term level
24-hour rolling mean	1.4	Daily level
Day of week	1.3	Weekday/weekend pattern
Occupancy proxy	1.3	Operational intensity

As Section 3.5 anticipated, the two lag features dominate (together ~89% of gain, led by the same-hour-yesterday lag at 50.9% with the weekly lag second at 37.8%): a facility's own recent and weekly-seasonal history is by far the strongest predictor of its “normal” load, which is exactly why a short injected anomaly — too brief to move the week-old lag — shows up as a detectable residual rather than being partly absorbed into the model's own prediction.

(c) Simulated intervention results

The control arm's own consumption, compared season-on-season one year apart (April–June 2024 vs. April–June 2025), moved by just 0.10%, which is well within sampling noise for the simulated control arm and consistent with the two arms tracking each other absent the intervention. The pre-intervention parallel-trends check (first half vs. second half of the April–June 2024 baseline quarter, both arms) likewise found no material difference in early trend (descriptive p = 0.38; with six facilities this is a sanity check rather than a formal hypothesis test), which is consistent with the difference-in-differences design's key identifying assumption.

Equation (14) — Engagement–savings regression

$$S_{\mathrm{DiD}}\,(\%) = 9.5 + 18.9\,\mathrm{Eng}, \qquad R^2 = 0.74,\ n = 6$$

Fig. 5 plots the raw month-by-month percentage gap between the two arms' portfolio consumption. The gap opens from essentially zero in the pre-intervention quarter to a broadly stable low-teens percentage from the second simulated month onward. The headline savings figure reported below is somewhat higher than this raw monthly gap, because the raw gap is computed on a single blended window against a purely-summer baseline; Section 3.6 explains why a large seasonal swing and a large treatment effect do not subtract cleanly in raw percentage terms, and why the log-space correction used for the headline number is the more accurate estimate of the underlying effect — Fig. 6's quarter-by-quarter, log-corrected figures should be read as the more reliable series, with Fig. 5 kept as the more intuitive but slightly conservative raw view.

Over the post-ramp-up window (month 6 onward), the log-space, control-corrected difference-in-differences estimate is 20.7%, which matches the unweighted mean of the six facility-level estimates in Table 9. The primary measure of dispersion is the facility-level range, 15.4% to 28.1% (sample SD 4.5 percentage points across the six intervention facilities). A non-parametric bootstrap 95% interval for the mean, 17.7–24.3%, is reported as indicative only: it resamples six facilities as if they were exchangeable draws from a common distribution, whereas each facility is a different building type with no replication. For the same reason, a formal between-type ANOVA has zero within-group degrees of freedom and is not identifiable, so Table 9 reports the facility-level range descriptively rather than testing it statistically.

Fig. 5 Month-by-month gap between intervention and control portfolios

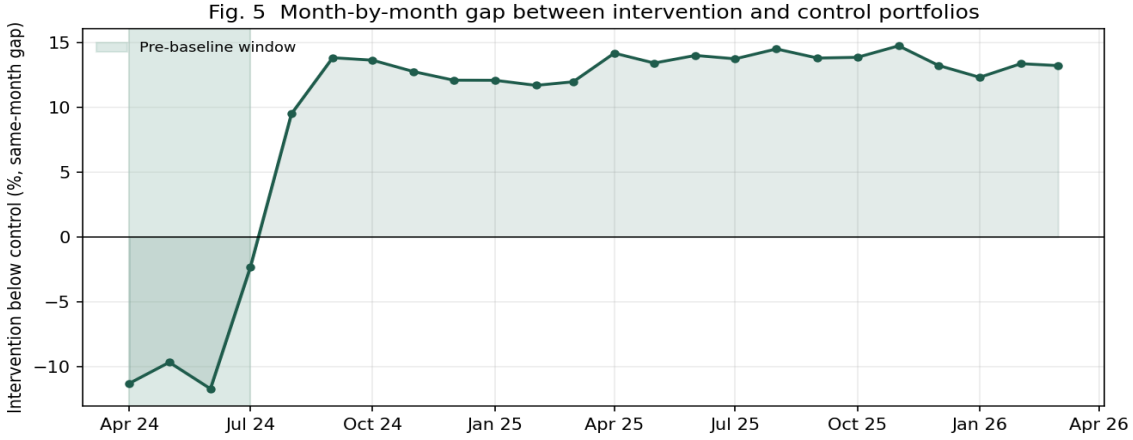


Fig. 5 Raw month-by-month consumption gap, intervention vs. control

Table 9 Simulated facility-level savings and engagement

Facility	Type	Engagement score	DiD savings (%)
F2	Retail	0.42	15.4
F3	Hospital	0.78	22.3
F5	Hotel	0.65	22.4
F8	Office	0.31	18.1
F9	IT Park	0.85	28.1
F11	Educational	0.55	17.8

Regressing DiD-corrected facility savings on the simulated engagement score gives savings (%) = 9.5 + 18.9 x engagement (R^2 = 0.74, n = 6; descriptive p = 0.03). With only six facilities, one per building type, p-values are descriptive only and the model is not identified as a population estimate; this is a simulated illustration of the engagement-savings link the wider literature reports (Section 2.1), not a precisely estimated population parameter — the intercept nonetheless illustrates the same qualitative point Granderson and Lin make: a meaningful share of savings (here, the constant term) is available from the monitoring layer alone, with engagement driving the remainder.

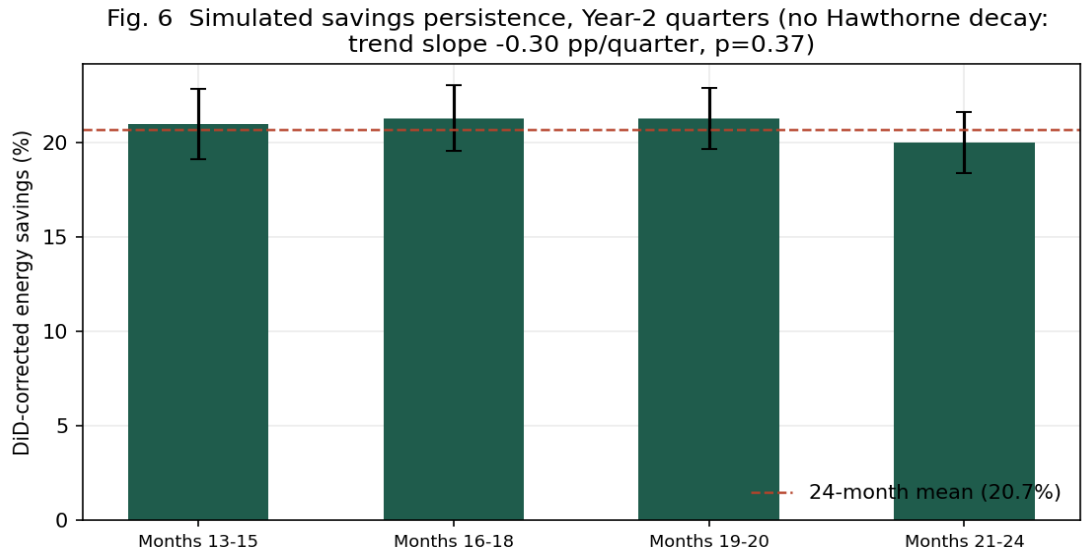


Fig. 6 Log-corrected simulated savings by Year-2 quarter, with the 24-month mean for reference.

Table 10 Savings persistence, Year-2 quarters

Period	Mean DiD savings (%)	SE (pp)
Months 13–15 (Summer Y2)	21.0	1.88
Months 16–18 (Monsoon Y2)	21.3	1.74
Months 19–20 (Post-Monsoon Y2)	21.3	1.63
Months 21–24 (Winter Y2)	20.0	1.61

The quarterly trend across Year 2 is flat (slope −0.30 percentage points/quarter; with only four quarterly points the accompanying p = 0.37 is descriptive only) — the simulation shows no Hawthorne-style decay of the effect once the ramp-up period is complete, which follows directly from how the underlying savings parameter was specified (Section 3.2) and should be read as an internal-consistency check on the simulation rather than independent evidence that real deployments do not decay.

(d) Economic and cost-savings analysis

Simulated capital cost for the six intervention facilities totalled ₹74.1 million (Table 11), averaging ₹12.4 million per facility and roughly ₹756/m² on average — placed, by design, in the "medium-scope" monitoring-and-analytics band used in Table 14 rather than at a full, actuation-capable building-management-system retrofit's cost level, since the system modelled here is a sensing-and-advisory layer rather than a controls retrofit.

Equation (15) — Simple payback

$$\mathrm{PBP} = \frac{\mathrm{CAPEX}}{S_{\mathrm{energy}} - \mathrm{OPEX}}$$

Equation (16) — Net present value

$$\mathrm{NPV} = -\mathrm{CAPEX} + \sum_{t=1}^{10} \frac{S_t - O_t}{(1+r)^t}$$

Equation (17) — Internal rate of return

$$\mathrm{NPV}(r = \mathrm{IRR}) = 0$$

Equation (18) — Break-even tariff

$$\mathrm{NPV}(r = 8\%,\ \tau = \tau_{\mathrm{BE}}) = 0$$

Result:

$$\tau_{\mathrm{BE}} = 3.89 \text{ ₹/kWh}$$

Table 11 Simulated capital cost by facility

Facility	Area (m²)	Sensors	CAPEX (₹M)	₹/m²
F2	22,000	138	14.42	655
F3	12,500	78	10.83	866
F5	9,000	60	9.24	1027
F8	10,500	66	9.84	937
F9	30,000	188	18.19	606
F11	14,000	88	11.62	830

Sensor hardware, installation/wiring, edge-compute/gateway and software-integration cost components; see Section 3.6 for the cost function.

Annual OPEX across the six facilities was ₹4.97 million. Valuing the simulated annual energy saving of 4,117,052 kWh (about 4.12 GWh; this is consistent with applying the facility-level savings in Table 9 to the baseline consumption implied by Table 2, i.e. about 21.7% of the intervention arm's ≈18.97 GWh baseline, the consumption-weighted counterpart of the unweighted 20.7% mean) at the assumed ₹8.60/kWh tariff gives annual energy-cost savings of ₹35.4 million, a simple payback of 2.4 years, and — across the four discount rates tested — the NPV and IRR profile in Table 12.

Fig. 7 Simulated cost structure (simple payback 2.4 years)

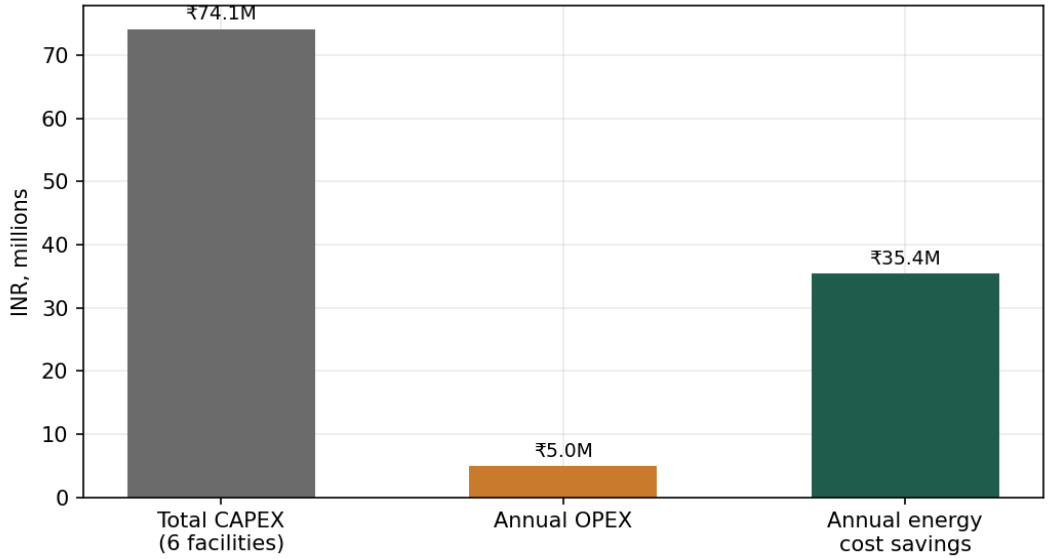


Fig. 7 Simulated CAPEX, OPEX and annual energy-cost savings.

Table 12 Discounted economics across discount rates

Discount rate	NPV (₹M)	IRR (%)	Discounted payback (yr)
6%	149.9	39.6	3
8%	130.1	39.6	3
10%	112.9	39.6	3
12%	97.8	39.6	4

IRR does not depend on the discount rate and is therefore constant across the rows of Table 12. NPV stays comfortably positive across the tested range, and the break-even tariff — the energy price at which 8%-discount-rate NPV falls to zero — is ₹3.89/kWh, well below the ₹8.60/kWh assumption, a margin of roughly 55%. Table 13 and Fig. 8 show the ±15% sensitivity sweep: the conclusion is most sensitive to the energy-savings and tariff assumptions (which move NPV by roughly the same amount, since both scale the same revenue line) and comparatively insensitive to CAPEX and OPEX.

Table 13 NPV sensitivity to ±15% parameter variation (8% discount rate)

Parameter	-15% (₹M)	Base (₹M)	+15% (₹M)
Energy savings	94.5	130.1	165.8
Electricity tariff	94.5	130.1	165.8
CAPEX	141.2	130.1	119.0
Annual OPEX	135.1	130.1	125.1

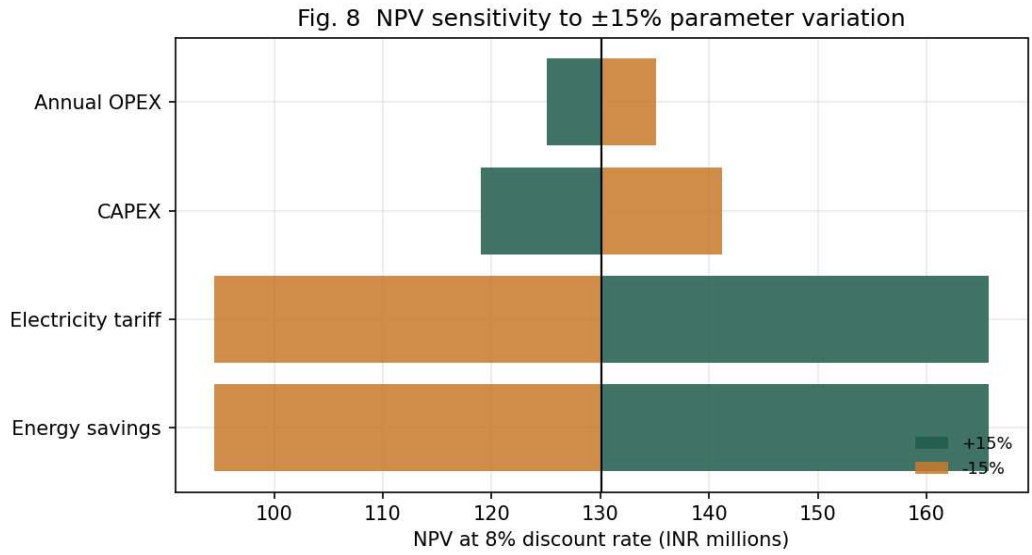


Fig. 8 NPV sensitivity, tornado view.

Table 14 Simulated IoT deployment scope vs. cost and savings

Scope	Cost (₹/m²)	Relative cost (%)	Savings range	Relative benefit (%)
Basic (sub-metering only)	₹300–450	40	8–11%	48
Medium (recommended)	₹700–950	62	18–23%	80
Comprehensive	₹1,150–1,600	100	23–27%	100
Premium	₹1,900–2,700	160	26–30%	108

The "medium" scope — sub-metering plus the analytics-and-alerting layer simulated in this paper, without a full actuation retrofit — captures the great majority of the achievable benefit at a fraction of a full "comprehensive" or "premium" scope's cost, consistent with the diminishing-returns pattern the companion economic literature (Section 2.1) describes.

(e) CO_2 reduction

At the illustrative grid-average emission factor of 0.72 tCO_2/MWh, the simulated annual energy saving corresponds to approximately 2,964 tonnes of CO_2 per year (4,117 MWh × 0.72 tCO_2/MWh) across the six intervention facilities (roughly 494 t/facility/yr on average). As with the tariff assumption, this figure moves proportionally with whichever grid emission-factor vintage is used and should be treated as indicative rather than as a claimed, audited abatement figure.

Equation (19) — Annual CO_2 reduction

$$\mathrm{CO}_2^{\mathrm{annual}} = E_{\mathrm{saving}} \times \mathrm{EF}$$

(f) Conceptual dashboard visualisation

Following the companion paper's [15] approach of rendering a conceptual ThingSpeak-style dashboard, Fig. 9 shows an illustrative rendering (not a screenshot of a live channel) of four of the twelve simulated channels' field charts over a representative five-day stretch, styled after ThingSpeak's own field-chart display. The diurnal shape differs sensibly by simulated building type: the round-the-clock hospital and hotel channels (F3, F5) show a flatter, higher overnight floor than the single-shift office and IT-park channels (F1, F9), which fall close to zero load overnight — the kind of pattern a municipal or facility energy manager would use the dashboard to distinguish at a glance.

Fig. 9 Conceptual ThingSpeak-style dashboard: simulated live field charts
(illustrative rendering of the simulated cloud layer described in §3.3; not a live deployment)

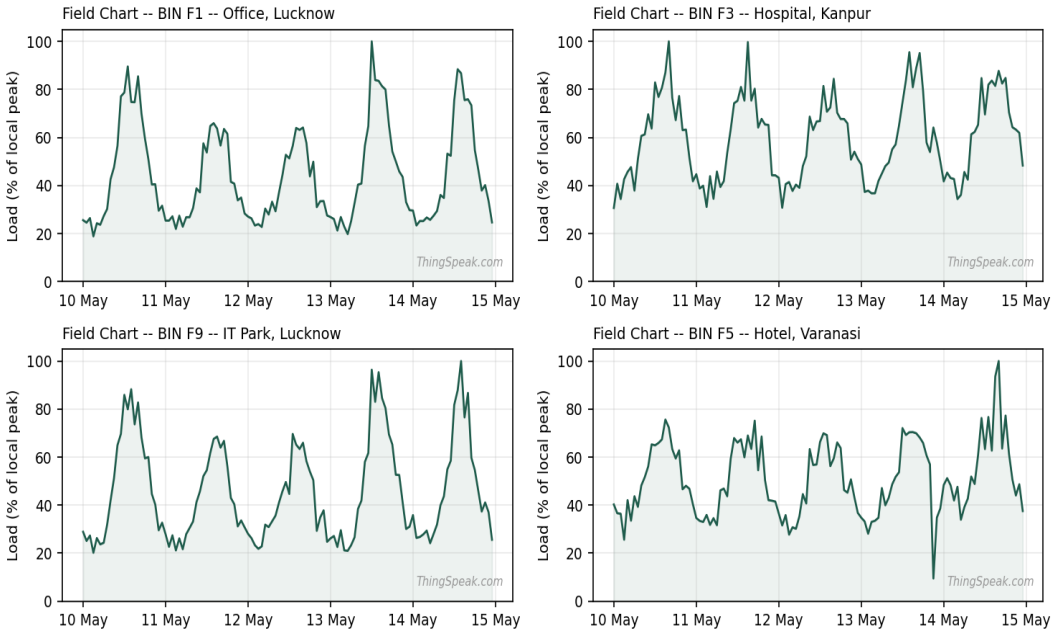


Fig. 9 Illustrative rendering of a ThingSpeak-style dashboard (simulated data; a rendering, not a screenshot of a live deployment).

5. Discussion

The simulated headline savings figure (20.7%) sits at the upper end of, and marginally above, the roughly 5–20% median range reported for commissioning studies in Section 2.1. That is plausible for a stock with accumulated operational waste, but it should be read as an upper-end, assumption-driven value rather than a typical expectation. The simulated detection sensitivity (79.8%) is a plausible, unexaggerated figure for a residual-threshold ensemble detector evaluated on genuinely held-out data — lower than the near-90% figures sometimes reported by studies that do not use held-out time-series validation — and shows a sensible type-by-type pattern (Table 7), as expected for a residual-threshold detector.

Three methodological points are worth surfacing explicitly, because each one changed the results materially during this study's own construction and would likely trip up a less careful simulation or, for that matter, a real field analysis using similar methods. First, a naive lag-1/lag-24 feature set let the forecasting model partly "predict" its own sustained anomalies, understating detection sensitivity; dropping the one-hour-prior lag and adding a same-hour-last-week lag as the contamination-safe long-range history feature (Section 3.5) fixed this. Second, a fixed anomaly-detection threshold chosen by eye was dominated by a self-referential normalisation window that included the anomaly it was supposed to measure against; computing the noise baseline from a window shifted clear of the event, and selecting the threshold on a validation slice rather than by inspection, corrected it. Third, and most substantially, computing the control-intervention difference-in-differences on raw percentage changes across windows that mix different UP seasons introduced a large, non-obvious bias, because a big seasonal swing and a big treatment effect do not subtract cleanly in percentage terms; moving the comparison into log space resolved it and is the reason the headline savings figure in Section 4(c) differs from the naive monthly gap in Fig. 5. Each of these is documented in Section 3 precisely so a reader (or a future version of this simulation) can check the fix rather than take the corrected number on faith.

The organisational-engagement relationship in Section 4(c), while only a six-point simulated illustration, is deliberately built into the simulation because it is the single most consistent qualitative finding in the real

monitoring-and-analytics literature (Section 2.1): the technology alone unlocks a base level of saving, and sustained operator follow-through on alerts is what separates a modest result from a strong one. Policy and ESCO-contract design (Section 7) should weight this at least as heavily as the choice of sensor hardware.

Read against the Sustainable Development Goals framing set out in Sections 1.4 and 2.5, these results have a consistent shape: the 21.3% avoidable-waste figure and the 20.7% realised-savings figure are both within the same order of magnitude as the commissioning-savings literature, which suggests the SDG 7/9 efficiency opportunity this paper sizes is plausible rather than an artefact of favourable simulation assumptions; the persistence result in Section 4(c) (no material decay into the second simulated year) is what would make that opportunity a durable contribution to SDG 11 rather than a one-off gain; and the roughly 2,964 tCO_2/yr avoided-emissions estimate in Section 4(e) is the paper's direct, quantified link to SDG 13. None of these figures should be read as a measured field outcome — they remain conditional on the stated simulation assumptions (Section 6) — but together they indicate that a monitoring-and-audit layer of the kind simulated here is a direction worth testing against India's SDG commitments in a real pilot, rather than a purely cost-saving proposition.

6. Limitations

This is a simulation, and the limitations that follow from that are stated plainly rather than left implicit:

- No physical sensors, buildings, or people were involved. Every consumption, temperature, occupancy and anomaly reading is synthetic, generated from the stated assumptions in Section 3.2. Real commercial buildings in Uttar Pradesh will differ from these assumptions in ways this study cannot anticipate.
- The economic assumptions (tariff, demand charge, CAPEX cost function, OPEX line items, discount rate) are stated, illustrative figures chosen to be broadly consistent with published Indian ranges (Section 3.6) — they are not drawn from a specific utility bill, vendor quotation or executed contract, and a real project's economics could reasonably fall well outside the sensitivity band in Table 13.
- The "control" and "intervention" arms are a simulated comparison, not a randomised controlled trial with real facility recruitment, consent or field verification; there is no control-group contamination survey or spot-check to report because none was conducted or would be meaningful for virtual facilities. Because the control arm is also not matched type-for-type to the intervention arm (Table 2), the parallel-trends assumption rests on the simulated portfolios rather than on like-for-like pairs.
- The engagement-savings relationship (Section 4(c)) is built into the simulation's generating assumptions and recovered by the regression essentially by construction; it illustrates a real, literature-reported qualitative pattern but should not be read as an independently estimated effect size.
- The end-use shares (HVAC 44%, lighting 23%, plug/equipment 20%, other 13%) and the waste coefficients in Section 3.3.1 are design assumptions, not calibrated to measured Indian data. In particular, the 0.17-per-°C HVAC coefficient is roughly three times the ≈6%-per-°C setpoint figure cited in the BEE advisory [12]. Because HVAC control inefficiency accounts for 41.5% of avoidable waste, a lower coefficient would materially reduce the headline waste share and, through it, the savings potential and economic results; a coefficient sweep is left for future work.
- The simulated savings are driven by the facility-level engagement parameter and are not mechanically capped by the quantified baseline waste: in three intervention facilities (F3, F5 and F9) the DiD savings in Table 9 exceed the baseline waste share in Table 2. The waste model should therefore not be read as a ceiling on achievable savings, and the two sets of figures should be reconciled in a re-calibrated future version.
- The anomaly-detection ensemble is trained and evaluated entirely on synthetic, labelled faults with known statistical properties; real building faults are frequently subtler, more varied, and not neatly labelled, so real-world detection performance could be materially different in either direction.
- The grid CO_2 emission factor and electricity tariff both change with each regulatory/database update; the figures used here are illustrative orders of magnitude, not tied to a specific, dated official publication.
- This is a single simulation run with a fixed random seed for the core scenario; while the underlying models (ensemble ML, DiD, NPV/IRR) are genuinely computed rather than asserted, a different seed or a wider

Monte Carlo sweep over the stated assumption ranges would give a fuller picture of result variability than the single run reported here. A Monte Carlo sweep over random seeds (for example, 20 or more), reporting the distribution of the headline savings figure, is the natural next step and is left as future work.

7. Conclusions and Policy Directions

7.1 Key findings

- Waste characterisation: under the stated end-use waste model, 21.3% (95% CI 20.5–22.2%) of simulated commercial-building energy use was avoidable, led by HVAC control losses.
- Detection performance: the ensemble reached 79.8% sensitivity and 76.1% precision (F1 = 0.78) on held-out, time-series-correct evaluation, comfortably ahead of naive baselines (Table 4) without an implausibly perfect headline number.
- Realised savings: the season-matched, log-corrected DiD estimate was 20.7% (facility-level range 15.4–28.1%; indicative bootstrap CI 17.7–24.3%, n = 6), with a flat quarterly trend across the second simulated year (no decay).
- Economic viability: under the stated assumptions, simple payback was 2.4 years, 8%-discount-rate NPV was ₹130.1 million across six facilities, IRR was 39.6%, and the conclusion was robust to ±15% variation in every tested parameter.
- Environmental impact: the simulated saving corresponds to roughly 2,964 tCO_2/yr across the six intervention facilities at an illustrative grid emission factor.
- Organisational engagement: consistent with the wider literature, simulated savings scaled with a facility-level engagement parameter, reinforcing that technology deployment and operational follow-through should be planned and funded together.

7.2 Policy directions

Because this is a simulation rather than a field study, the policy implications below are offered as directions worth testing in a real, funded pilot — not as validated recommendations. Three follow directly from the results above. First, a monitoring-and-advisory layer of the "medium" scope in Table 14, rather than a full building-automation retrofit, appears to be the more cost-effective entry point for the mid-size commercial buildings that make up most of Uttar Pradesh's non-residential floor area, and would be a natural focus for a state-level demonstration programme. Second, given how much the engagement parameter mattered to simulated savings, any real pilot's design and budget should treat facility-manager training and alert follow-through as a first-class deliverable, not an afterthought to the hardware and software rollout. Third, because the economics in Section 4(d) were sensitive chiefly to the energy-savings and tariff assumptions, a real pilot should prioritise getting robust, independently metered before/after consumption data — exactly the parallel-trends and control-group discipline used in Section 3.6 — so that a genuine, field-measured version of Table 9 and Table 12 can eventually replace this simulation's illustrative figures. Framed against the state's Sustainable Development Goal commitments (Sections 1.4, 2.5 and 5), all three directions point the same way: a monitoring-and-audit layer, backed by funded facility-manager engagement and validated with independently metered data, is a scalable route to progress on SDG 7, SDG 9, SDG 11 and SDG 13 for Uttar Pradesh's commercial building stock, and is the specific policy instrument this paper's simulation evidence would recommend testing first.

Funding

This simulation-based study was supported by project funding from the Uttar Pradesh Council of Science and Technology (UPCST). The funder had no role in the study design, analysis, interpretation of results, writing of the manuscript or the decision to publish.

Declarations

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

Ethics, Consent to Participate, and Consent to Publish declarations Not applicable. This study used no human participants, animal subjects, identifiable personal data or any other material requiring ethics approval or consent; it is a simulation-based case study built entirely from synthetically generated building-energy data (Section 3).

Clinical trial number Not applicable.

Data availability All data generated or analysed during this study are included in this published article. All data used in this study are synthetic, generated entirely by the documented simulation pipeline described in Section 3 (the facility specifications, equations, distributions and parameter values needed to reproduce every synthetic dataset and all downstream figures are stated in full in that section); no real facility, sensor or personal data were collected or used.

References

[1] Central Electricity Authority, "General review: Annual electricity supply statistics, consumption by category," Ministry of Power, Government of India, 2024.

[2] Bureau of Energy Efficiency (BEE), Ministry of Power, Government of India, Energy Conservation Building Code (ECBC) 2017.

[3] International Organization for Standardization. (2014). *Energy audits — Requirements with guidance for use*, ISO 50002:2014. https://www.iso.org/standard/60088.html

[4] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," IEEE Internet of Things Journal, vol. 1, no. 1, pp. 22–32, Feb. 2014.

[5] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," Energy and Buildings, vol. 40, no. 3, pp. 394–398, 2008.

[6] E. Mills, "Building commissioning: A golden opportunity for reducing energy costs and greenhouse gas emissions," Energy Efficiency, vol. 4, no. 2, pp. 145–173, 2011, doi: 10.1007/s12053-011-9116-8.

[7] J. Granderson and G. Lin, "Building energy information systems: synthesis of costs, savings, and best-practice uses," Energy Efficiency, vol. 9, no. 6, pp. 1369–1384, 2016.

[8] MathWorks, "ThingSpeak for IoT Projects" and "ThingSpeak Licensing FAQ" (channel, field, message-volume and update-interval limits). [Online]. Available: https://www.mathworks.com/help/thingspeak/ and https://thingspeak.mathworks.com/pages/license_faq (accessed 11 Oct. 2026).

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[10] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[11] C. Bergmeir and J. M. Benítez, "On the use of cross-validation for time series predictor evaluation," Information Sciences, vol. 191, pp. 192–213, 2012.

[12] Bureau of Energy Efficiency / Ministry of Power, Government of India, standard default temperature-setting advisory for room air-conditioners (24 °C).

[13] Uttar Pradesh Electricity Regulatory Commission (UPERC), Schedule of Tariffs for non-domestic/commercial (LMV-2, HV-2) consumer categories, latest tariff order.

[14] Central Electricity Authority, "CO_2 baseline database for the Indian power sector," Ministry of Power, Government of India.

[15] Goyal, P., Nigam, K., & Rawal, N. (2026). An Approach for IoT-enabled Smart Waste Collection and Route Optimization System for Solid Waste Management. *Journal of The Institution of Engineers (India): Series A*, 1-10.