An IIoT-enabled real-time ESG carbon intelligence and MRV framework for industrial facilities: from sensor-level emission measurement to auditable carbon-credit quantification

[1]Saswat Swain [2] Chandrasekhar Panda [3]Pravat Satpathy [4]Sonia Mohapatra

[1,2]Ph.D Scholar, [3,4]R & D Division

[1,2]KSOM - KIIT Deemed to be University ,[3,4]Tekstein Solutions,India

[1]saswat.mck@gmail.com [2]contact@chandrasekarpanda.com [3]pravat@teksteinsolutions.com [4]sonia@teksteinsolutions.com

Abstract—Corporate greenhouse-gas (GHG) inventories for manufacturing sites are still assembled from monthly utility invoices and engineering estimates, which limits both the granularity and the timeliness of environmental, social and governance (ESG) disclosure. This paper develops and evaluates an Industrial Internet of Things (IIoT) architecture that couples 15-minute machine-level electrical and thermal metering to a GHG accounting engine, a measurement-reporting-verification (MRV) layer and a conservative carbon-credit quantification module. Two contributions are formalised: a provenance-tier traceability index that propagates data quality into the reported inventory, and a deduction rule that links the creditable share of an emission reduction to both measurement uncertainty and traceability. The framework is evaluated on a physics-based one-year simulation of a twelve-machine, three-line discrete manufacturing plant (420,480 machine-interval records) containing degradation, telemetry loss, grid outages, injected faults and on-site photovoltaic generation. Against conventional periodic reporting, continuous monitoring reduced machine-level attribution error from 11.29 percent to 0.33 percent and reporting latency from a median of 90.6 days to 8.0 minutes, both significant under paired Wilcoxon tests. Three findings qualify the benefit. First, the annual facility-level inventory improved only marginally, because emission-factor uncertainty contributes over 92 percent of the total variance and is unaffected by better activity data. Second, unsupervised outlier detection on raw telemetry was close to useless (F1 = 0.121); the same algorithm on context-conditioned residuals reached F1 = 0.340, indicating that feature construction rather than model complexity governs performance. Third, day-ahead carbon-intensity forecasting failed for every model tested. The quantified benefit of granularity appears instead in the credit module, where a daily rather than monthly adjusted baseline reduced the conservative deduction from 14.33 percent to 6.24 percent, raising the creditable quantity by 21.8 tCO2e. All results are simulated; no hardware measurements are claimed.

Index Terms—Industrial Internet of Things, ESG reporting, GHG accounting, Scope 1 and Scope 2 emissions, carbon MRV, carbon-credit quantification, edge computing, data traceability, anomaly detection.

I. INTRODUCTION

Manufacturing accounts for a large share of global energy-related greenhouse-gas (GHG) emissions, and disclosure of those emissions is moving from voluntary practice to regulated obligation. Under the GHG Protocol Corporate Standard [1] a facility must report direct combustion emissions (Scope 1) and emissions from purchased electricity (Scope 2). The IFRS S2 climate disclosure standard [2], the Global Reporting Initiative disclosure 305 [3], and, in India, the Business Responsibility and Sustainability Report [4] all now require these figures alongside intensity metrics normalised to output.

In practice the underlying data are thin. A typical plant assembles its annual inventory from twelve electricity invoices, twelve gas meter readings and a fuel stock reconciliation. Machine-level figures, when reported at all, are produced by multiplying nameplate ratings by assumed operating hours. The resulting inventory is a small number of measurements stretched across a large reporting matrix, and it arrives months after the emissions occurred. Reviews of carbon accounting practice identify exactly this gap between accounting frameworks and measurement infrastructure [5], [6].

Industrial Internet of Things (IIoT) instrumentation is an obvious remedy, and a growing literature proposes IoT-based carbon monitoring [7], electricity-data-driven emission estimation [8] and digital measurement, reporting and verification (MRV) architectures [9]. Two weaknesses recur. First, most studies report an architecture and a dashboard but do not quantify what the additional data buys relative to the conventional practice it replaces. Second, data provenance is treated qualitatively: papers assert that continuous data are more auditable without defining a measure of auditability that could enter a reported figure or a credit calculation.

This paper addresses both. We construct a physics-based simulation of a discrete manufacturing plant, implement conventional periodic reporting and continuous IIoT reporting over identical ground truth, and measure the difference. The specific contributions are:

1) A traceability index τ that assigns every reported emission increment to a provenance tier and aggregates tier weights by emission mass, so that data quality is carried in the inventory rather than in a footnote.

2) A conservative credit-deduction rule that links the creditable share of a verified emission reduction jointly to its Monte-Carlo uncertainty and to τ, and a demonstration that the temporal resolution of the adjusted baseline, not the precision of the meters, dominates that deduction.

3) A gap-length-stratified reconstruction estimator for telemetry loss, with the finding that the estimator ranking reverses across gap length and that the common zero-fill default introduces a systematic annual under-report.

4) A paired experimental comparison of periodic and continuous reporting across latency, completeness, attribution accuracy, fault detection, prediction, abatement and uncertainty, including three results where the continuous system does not help.

Section II positions the work, Sections III to V give the architecture, formulation and simulation, Sections VI and VII give the experiments and results, and Sections VIII to X discuss implications and limits.

II. RELATED WORK AND RESEARCH GAP

Work relevant to this paper falls into four groups. Carbon accounting methodology defines what must be computed: emission factors, activity data and scope boundaries are set out in the IPCC inventory guidelines [10] and operationalised for organisations by ISO 14064-1 [11]. Electricity-sector accounting reviews [5] show that indirect emission estimation is dominated by emission-factor choice rather than by metering. Digital MRV research [9] proposes IoT, remote sensing and distributed ledgers as a route to verifiable emission data, mostly in land-use contexts, and must ultimately satisfy the verification requirements of ISO 14064-3 [12]. Industrial IIoT monitoring studies [7], [13] demonstrate sensing architectures for real-time CO_2 and energy visibility, and machine-learning studies estimate emissions from electricity data [8].

Table I positions representative studies against the capabilities required for an auditable industrial ESG-MRV pipeline. The recurring gap is the right-hand columns: quantified comparison against the incumbent reporting method, an explicit data-quality measure entering the reported number, and propagation of that measure into credit quantification.

TABLE I. POSITIONING AGAINST REPRESENTATIVE PRIOR WORK

Study focus and data basis	**Machine level**	**Quantified vs. manual**	**Provenance metric**	**Credit link**
Electricity carbon accounting review, factors and statistics [5]	No	No	No	Partial
Emission estimation from grid meter data [8]	No	No	No	No
IIoT real-time CO_2 monitoring architecture [7]	Partial	No	No	No
Continuous emission metering (CEMS) review [13]	No	No	Partial	No
Digital MRV protocol review, IoT and ledger [9]	No	No	Qualitative	Partial
This work: 15-min IIoT plus fiscal meters	Yes	Yes, paired tests	Yes, τ index	Yes, δ rule

Accordingly the research question is whether continuous IIoT-derived operational data measurably improves the accuracy, timeliness, completeness and traceability of industrial ESG carbon reporting relative to conventional periodic reporting, and where it does not.

III. SYSTEM ARCHITECTURE AND DATA MODEL

Fig. 1 shows the eight-layer architecture. Sensing (L1) uses split-core current transformers and a three-phase power-quality front end per machine, plus PT100 thermocouples, capacitive humidity sensors and non-dispersive infrared CO_2 sensors for the environmental channel. An ESP32-S3 node (L2) samples at 1 Hz, aggregates to 15-minute active-energy, mean power-factor and mean-current records, and buffers to flash so that a transport outage delays rather than destroys a record. Transport (L3) is MQTT over TLS in the plant with a LoRaWAN fallback. The Raspberry Pi gateway (L4) applies the validation rules and reconstruction of Section IV at the edge, so that a record reaches the database already tagged with a provenance tier.

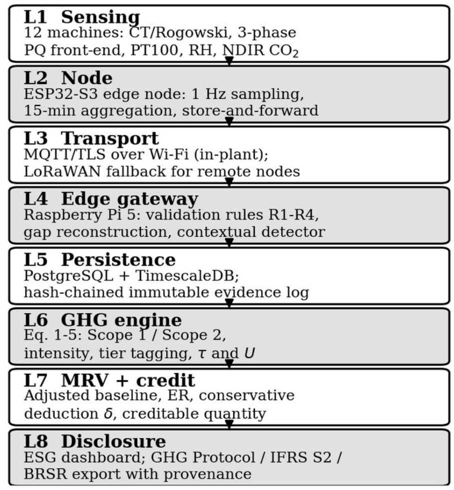

Fig. 1. Eight-layer IIoT-ESG-MRV reference architecture. Layers L1–L4 are edge-resident; L5–L8 are server-side. Each layer emits provenance metadata consumed by the traceability index of Eq. 5.

Persistence (L5) is PostgreSQL with a time-series extension. Each 15-minute record is appended to a hash-chained evidence log: the digest of record k includes the digest of record k−1, so that a retrospective edit is detectable without a distributed ledger. This is an integrity control, not an assurance mechanism; the distinction matters and is returned to in Section IX.

Layers L6 to L8 compute the inventory, the MRV quantities and the disclosure export. Two data classes are kept strictly separate throughout. Activity data (kWh, GJ, litres) drive the GHG calculation. Environmental data, including indoor CO_2 concentration, are treated only as anomaly indicators and corroborating evidence. Indoor CO_2 concentration in parts per million is not a corporate emission quantity, and conflating the two is a modelling error that appears in parts of the IoT carbon-monitoring literature.

IV. MATHEMATICAL FORMULATION

A. GHG accounting

Machine i consumes energy over interval Δt at active power P. Interval and monthly energy are

$$E_{i,k} = P_{i,k} \cdot \Delta t, \quad E_i = \Sigma_k E_{i,k} \qquad (1)$$

with $\Delta t = 0.25$ h. Scope 2 emissions follow the location-based method with grid emission factor EF_g, and on-site self-consumed photovoltaic energy E_{pv} is excluded from grid import rather than credited separately:

$$C_{s_2} = (E_{grid}) \cdot EF_g / 1000, \quad E_{grid} = E_{site} - E_{pv} \qquad (2)$$

Scope 1 combines piped natural gas of energy content G and standby diesel volume V:

$$C_{s_1} = (G \cdot EF_{ng} + V \cdot EF_d) / 1000 \qquad (3)$$

Emission factors are documented defaults, not fitted values: EF_g = 0.716 $kgCO_2e/kWh$ from the national grid baseline database [14]; EF_{ng} = 56.157 $kgCO_2e/GJ$ and EF_d = 2.686 $kgCO_2e/L$ derived from IPCC default carbon contents [10] with AR6 100-year global warming potentials for CH_4 and N_2O. Carbon and energy intensity normalised to saleable output Q are

$$CI = (C_{s_1} + C_{s_2}) / Q, \quad EI = E_{site} / Q \qquad (4)$$

B. Provenance tiers and the traceability index

Every reported increment is assigned a tier $t \in \{1,2,3,4\}$ with weight $w(t) = \{1.00, 0.80, 0.50, 0.20\}$. Tier 1 is received and passes all validation rules; tier 2 is received but fails at least one;

tier 3 is reconstructed from a fitted model; tier 4 is a default or nameplate proxy. Four automated rules run at the edge: R1 plausibility, $0 \le P \le 1.15\, P\_rated$; R2 power factor within [0.45, 1.0]; R3 hourly reconciliation of summed node energy against the fiscal meter within 4 percent; R4 state consistency, a machine reported as running must draw more than 1.05 times its idle power. The traceability index is the energy-weighted mean tier weight,

$$\tau = \Sigma_k\, w(t_k)\, E_k \,/\, \Sigma_k\, E_k\,, \quad \tau \in [0.20, 1.00] \qquad (5)$$

τ is reported alongside the inventory. It is deliberately mass-weighted: a gap during a high-load interval degrades τ more than a gap at night, which is the behaviour an assurance provider would want.

C. Gap reconstruction

For a missing record the context-conditioned estimator uses the mean of received records for the same machine, hour-of-day and day type within a ±14-day window,

$$\hat{E}_{i,k} = \mathrm{mean}\{\, E_{i,j} : c(j) = c(k),\ |t_j - t_k| \le 14\ \mathrm{d} \,\} \qquad (6)$$

where $c(\cdot)$ is the context key. The alternatives evaluated are zero-fill, linear interpolation, and a hybrid that applies interpolation to gaps of at most 20 samples and the context estimator beyond.

D. Uncertainty and credit quantification

The inventory uncertainty is propagated by Monte-Carlo simulation with 10,000 draws consistent with the GUM approach [15], sampling emission factors and activity data independently from normal distributions whose 95 percent half-widths are the documented relative uncertainties. The baseline is production-adjusted by ordinary least squares on periods of length h,

$$\hat{C}\_base = a + b \cdot Q\,, \quad ER = \hat{C}\_base - C\_project \qquad (7)$$

The standard error of the annual baseline prediction is obtained by a moving-block bootstrap with block length seven periods, which respects the serial correlation of industrial load. Only reductions that survive a conservative deduction are treated as potentially creditable:

$$\delta = \min[0.6,\ \kappa \cdot \max(0, U\_rel - U_0) + \lambda(1 - \tau)] \qquad (8)$$

$$Cr = ER \cdot (1 - \delta) \qquad (9)$$

with $\kappa = 1$, $\lambda = 0.25$ and $U_0 = 0.05$. These parameters are proposed, not standardised; they follow the spirit of conservative-factor practice in existing methodologies [16] but are not drawn from any codified rule. Cr is a potential quantity computed from simulated data. It is not an issued credit, and no claim of additionality, permanence or eligibility is made.

V. SIMULATION ENVIRONMENT AND DATASET

The facility comprises twelve machines on three lines: computer-numerical-control machining, a thermal and forming line with a gas-fired heat-treat furnace and two injection moulders, and an assembly and finishing line with a gas-fired paint booth. A shared air compressor, common services, a 300 kWp rooftop photovoltaic array and a standby diesel generator complete the site. Sampling is at 15 minutes for one calendar year, giving 35,040 site records and 420,480 machine-interval records.

Machine power follows $P = [P\_idle + (P\_rated - P\_idle)\cdot \ell]\cdot \eta\_deg$, where cycle load ℓ is Beta-distributed with a shape parameter driven by an order-book demand series and η_deg grows with time since maintenance. Power factor falls with reduced loading, which drives distribution losses modelled as proportional to the inverse square of power factor. Furnace cold starts carry a physically derived thermal charge computed from refractory mass, specific heat, temperature rise and burner efficiency, so that batch scheduling has a mechanistic rather than assumed effect on gas consumption. Exogenous drivers, namely demand, ambient conditions, a clear-sky solar model with a stochastic clearness index, grid outages and network availability, are drawn once and shared between the baseline and intervention runs, making that comparison strictly paired.

TABLE II. PRINCIPAL SIMULATION PARAMETERS AND ANNUAL GROUND TRUTH

Quantity	Value	Unit
Machines / lines / horizon	12 / 3 / 365	– / – / d
Sampling interval	15	min
Machine-interval records	420,480	–
Telemetry loss (burst model)	2.64	%
Grid outage duration	36.2	h
Injected fault prevalence	4.40	%
Site electricity	1,272,725	kWh
Grid import	938,350	kWh
PV generation / self-consumed	380,400 / 86.3	kWh / %
Natural gas	6,664	GJ
Standby diesel	1,713	L
Saleable output	1,197,056	parts
Scope 1 / Scope 2 / total	378.8 / 671.9 / 1050.7	tCO_2e
Carbon intensity	877.7	kg/1000 parts
Random seed	20260416	–

Realistic phenomena are injected with known ground truth: two-shift weekday and single-shift Saturday operation with ten public holidays, seasonal and autoregressive demand variation, monsoon suppression of the solar clearness index, machine degradation between maintenance events, Class 0.5S metering error [17] with per-node calibration bias and drift, a two-state burst model for MQTT and LoRaWAN packet loss, grid outages that transfer load to the diesel generator, and five labelled fault classes: off-shift idle running, bearing wear, compressed-air leakage, heater element fault and power-factor collapse.

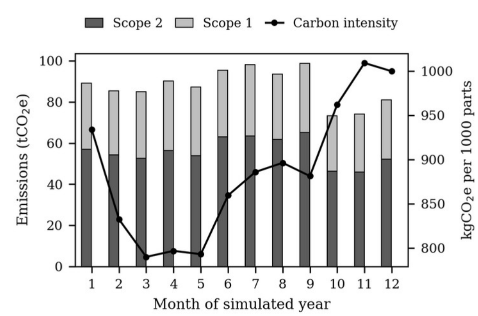

Fig. 2. Monthly Scope 1 and Scope 2 emissions and production-normalised carbon intensity (ground truth). Intensity falls in high-volume months because a substantial fixed load is amortised over more output.

A. Reporting scenarios

Scenario A reproduces conventional practice. Electricity is taken from the utility invoice on a billing cycle ending on the fifth of each month, so the reported month is offset from the calendar month; one month is an unread estimate trued up in the following period. Natural gas follows the same cycle. Diesel is reconciled by inventory difference on a 2000 L tank with a ±3 percent sight-gauge error. Machine-level figures are allocated by nameplate rating multiplied by scheduled hours and normalised to the invoice after an assumed ten percent common-services deduction. Scenario B reads the same physical fiscal instruments at 15-minute resolution on calendar boundaries, adds machine sub-metering, and applies the validation, reconstruction and tier assignment of Section IV. The systematic calibration bias of the fiscal meters is therefore common to both scenarios and cannot explain any difference between them.

VI. EXPERIMENTAL DESIGN AND HYPOTHESES

Ten experiments were run: reporting latency (E1), data completeness (E2), attribution accuracy (E3), fault detection (E4), emission prediction (E5), abatement and credit quantification (E6), telemetry-loss robustness (E7), emission-factor sensitivity (E8), production-volume sensitivity (E9) and renewable integration (E10). Five hypotheses are tested. H1: continuous data improve completeness. H2: continuous reporting reduces latency. H3: machine-level operational data improve attribution accuracy. H4: contextual features improve fault detection over raw-telemetry outlier detection. H5: continuous monitoring raises the traceability index.

Paired comparisons use the Wilcoxon signed-rank test, chosen because the paired differences are strongly non-normal and the sample sizes (twelve months or twelve machines) are too small for a reliable normality assessment; the matched-pairs rank-biserial correlation is reported as effect size. Machine-learning tasks use a strictly chronological split with no shuffling: for fault detection, January to August for training and September to December for testing; for prediction, 236 training days, 61 validation days and the remainder for testing. No feature uses information from the prediction horizon.

VII. RESULTS

A. Latency, completeness and traceability

Under Scenario A an emission increment becomes available in the published register a median of 2175.1 h (90.6 days) after it occurs, with a 95th percentile of 3160.5 h, driven by the invoice lag, internal consolidation and the quarterly review cycle. Under Scenario B the median is 0.134 h (8.0 minutes) and the mean 0.290 h, the mean exceeding the median because records lost to a network outage are committed only on recovery. The paired monthly medians differ significantly ($W = 0$, $p = 4.88\times10^{-4}$), a ratio of about 1.6×10^4. H2 is supported.

Completeness was assessed on the machine-by-month reporting matrix of 144 cells. Scenario A populates zero cells with metered evidence, resting on 36 primary instrument readings per year; Scenario B populates all 144 from 409,364 received observations at a 97.36 percent receipt rate ($W = 0$, $p = 4.88\times10^{-4}$). H1 is supported. Validation rule flag rates were 2.64 percent for R1, 0 percent for R2, 4.73 percent for R3 and 0 percent for R4; R1 exceedances concentrate in intervals containing injected wear and leak faults, so the plausibility rule doubles as a coarse fault screen. The resulting tier distribution by energy was 96.48 percent tier 1, 0.60 percent tier 2 and 2.91 percent tier 3, with no tier 4 records, giving $\tau = 0.984$ at attribution level against 0.200 for nameplate allocation. At whole-inventory level τ rises more modestly, from 0.746 to 0.896, because Scope 1 gas and diesel remain partly calculated in both scenarios. H5 is supported with that qualification.

B. Attribution accuracy

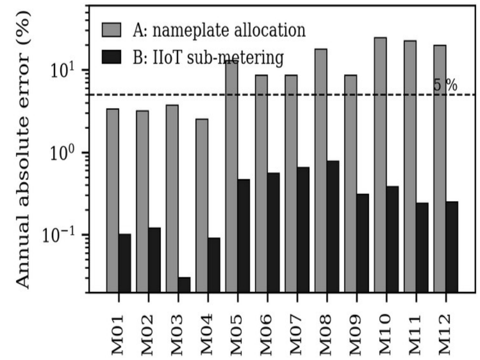

Fig. 3. Annual absolute error in machine-level electricity attribution, logarithmic axis. Nameplate allocation errs most on machines whose duty cycle departs from their nameplate rating: the air compressor (M08), the paint booth (M10) and the light assembly and packaging lines (M11, M12).

Mean absolute annual attribution error was 11.29 percent under Scenario A and 0.33 percent under Scenario B, with worst-case machine errors of 24.36 and 0.78 percent respectively ($W = 0$, $p = 4.88\times10^{-4}$, rank-biserial 1.00). At machine-month granularity the corresponding figures are 13.87 and 0.46 percent. H3 is supported at machine level.

At facility level the picture is different and deserves emphasis. Monthly facility totals were in error by 5.56 percent under Scenario A, almost entirely from billing-period misalignment, against 0.25 percent under Scenario B. But the annual facility total was in error by only +0.31 percent under Scenario A against +0.25 percent under Scenario B, because the billing offsets largely cancel over a full year and both scenarios read the same fiscal meter. Continuous monitoring therefore does not materially improve the headline annual number. Its benefit lies in the monthly and machine-level resolution beneath it.

C. Robustness to telemetry loss

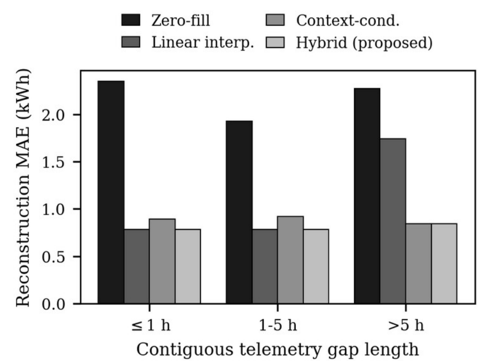

Fig. 4. Reconstruction error by contiguous gap length. The ranking of estimators reverses: interpolation is preferable for short gaps and the context-conditioned estimator for gaps beyond five hours.

Zero-fill, the common default when a record is absent, produced a −2.59 percent systematic annual energy under-report at the observed 2.64 percent loss rate. Over all held-out gaps the context-conditioned estimator gave the lowest error of the three single methods (MAE 0.873 kWh against 1.320 for interpolation and 2.210 for zero-fill), but the stratified view shows why a single method is the wrong choice: interpolation is better for gaps up to five hours (MAE 0.783 against 0.895 at one hour), while beyond five hours it degrades sharply and acquires a +24.5 percent bias as it bridges shift boundaries, where the context estimator holds at MAE 0.844. The hybrid rule attains the best overall error (MAE 0.818).

D. Fault detection

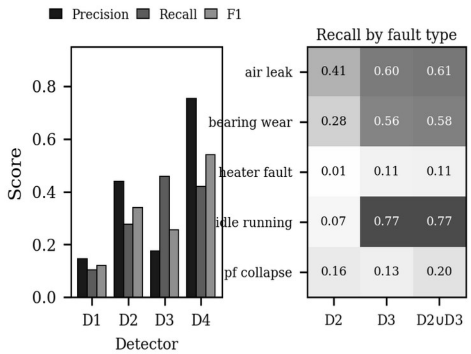

Fig. 5. Detector performance on the chronologically held-out test period (136,590 records, 6.72 percent prevalence). D1: Isolation Forest on raw telemetry. D2: Isolation Forest on context-conditioned residuals. D3: physical rule. D4: supervised random forest on the same contextual features.

Isolation Forest on raw telemetry (power, power factor, current, load) achieved F1 = 0.121 with ROC-AUC 0.567, barely better than chance. The identical algorithm on context-conditioned residual features reached F1 = 0.340 and ROC-AUC 0.699, a 2.8-fold gain in F1 from feature construction alone. A supervised random forest on the same features reached F1 = 0.540 and ROC-AUC 0.831 but requires labelled faults, which are rarely available in practice. H4 is supported, with the practical implication that effort belongs in residual modelling rather than in detector complexity.

Per-fault recall exposes the limits. Off-shift idle running is detected almost entirely by the physical rule (recall 0.768) and almost not at all by the residual detector (0.072), because an idling machine draws exactly its expected idle power and only the schedule context makes it anomalous. Air leakage and bearing wear are detected by both (union recall 0.61 and 0.58). Heater faults are essentially missed by every electrical detector (union recall 0.11), because the fault manifests in gas consumption; detecting it requires gas-side instrumentation that the electrical channel cannot substitute for. This is a boundary of the architecture, not a tuning failure.

E. Emission prediction

Day-ahead facility CO_2e was predicted from lagged emissions, planned production, scheduled hours and weather. Gradient boosting performed best (MAE 0.359 tCO_2e/day, R^2 = 0.845 against a test mean of 2.54 tCO_2e/day), with random forest close behind and ordinary least squares materially worse. Persistence failed because non-productive days drive the target near zero; for the same reason MAPE is unstable here and MAE is the primary metric.

TABLE III. DAY-AHEAD PREDICTION PERFORMANCE ON THE HELD-OUT TEST PERIOD

Model	**MAE**	**RMSE**	**MAPE (%)**	**R^2**
Facility CO_2e (tCO_2e/day), n = 61				
Persistence	1.058	1.510	163.4	−0.721
Linear regression	0.581	0.659	30.3	0.672
Random forest	0.409	0.495	15.0	0.815
Gradient boosting	0.359	0.453	13.8	0.845
Carbon intensity (kg/1000 parts), n = 51				
Persistence	92.96	133.03	8.97	−0.755
Linear regression	78.44	100.16	7.39	0.005
Random forest	96.16	115.81	9.12	−0.330
Gradient boosting	84.70	103.72	8.09	−0.067

Day-ahead carbon intensity, by contrast, could not be predicted by any model: the best R^2 was 0.005 and the tree ensembles were negative, meaning they underperformed the test-set mean. Intensity is a ratio of two strongly correlated quantities whose common variation cancels, leaving a residual that is close to noise at daily resolution. The level is predictable, the deviation is not. Claims in the literature that continuous data enable accurate intensity forecasting should be treated with caution at this granularity.

F. Abatement and credit quantification

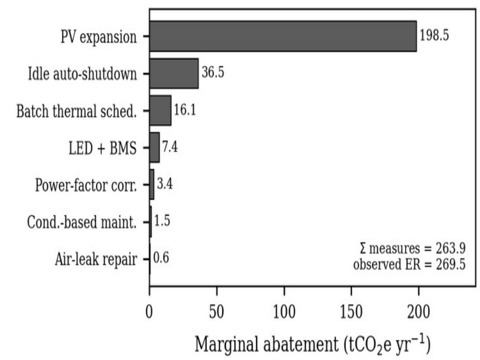

Fig. 6. Marginal abatement by measure, obtained by leave-one-out ablation of the intervention run. The residual between the sum of marginal effects and the observed reduction is the measure interaction term.

Six measures were applied in a paired intervention run: 30-minute idle auto-shutdown, condition-based maintenance, power-factor correction, compressed-air leak repair, consolidated furnace batching with a shortened hot-hold, LED and building-management retrofit, and photovoltaic expansion from 300 to 650 kWp. Site electricity fell from 1272.7 to 1189.6 MWh (−6.5 percent) and total emissions from 1050.7 to 781.2 tCO_2e (−25.65 percent), at essentially unchanged output (−0.14 percent); carbon intensity fell from 877.7 to 653.5 kg per 1000 parts. Monthly totals differ significantly ($W = 0$, $p = 4.88\times10^{-4}$).

Each reduction is traced to its mechanism by leave-one-out ablation rather than assumed. Photovoltaic expansion contributes 198.5 tCO_2e, idle shutdown 36.5, batch thermal scheduling 16.1 (entirely in gas, through avoided furnace cold starts), lighting and controls 7.4, power-factor correction 3.4, condition-based maintenance 1.5 and leak repair 0.6; the marginal effects sum to 263.9 against an observed 269.5, the 5.6 tCO_2e residual being interaction. Photovoltaic expansion shows clear diminishing returns: self-consumption falls from 86.3 percent of generation at 300 kWp to 72.6 percent at 650 kWp, and the marginal utilisation of the added capacity is only about 61 percent because surplus generation coincides with non-productive Sundays and holidays. Power-factor correction, often promoted as a carbon measure, delivers only 3.4 tCO_2e, since it acts solely on distribution losses of roughly 1.6 percent of site load.

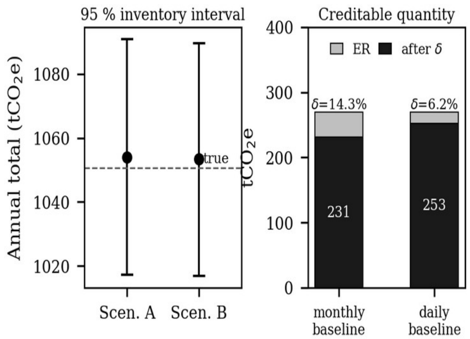

Fig. 7. Left: 95 percent Monte-Carlo interval for the annual inventory under both scenarios, against ground truth. Right: emission reduction and creditable quantity after the conservative deduction, for a monthly and a daily adjusted baseline.

The credit module exposes where granularity actually pays. The emission reduction is 269.5 tCO_2e by both direct difference and regression adjustment. Its uncertainty, however, depends on the resolution at which the adjusted baseline can be fitted. A monthly baseline (n = 12, $R^2 = 0.747$) gives a bootstrap standard error of 16.66 tCO_2e on the annual prediction and a total relative uncertainty on ER of 12.97 percent; a daily baseline (n = 365, $R^2 = 0.959$), which only continuous data can support, halves that standard error to 9.94 tCO_2e and reduces the relative uncertainty to 8.63 percent. Through Eq. 8 the deduction falls from 14.33 to 6.24 percent and the creditable quantity rises from 230.9 to 252.7 tCO_2e, a gain of 21.8 tCO_2e. Notably, the activity-data component of ER uncertainty is only 4.67 percent: the baseline model, not the metering, is the binding constraint.

G. Uncertainty and sensitivity

Monte-Carlo propagation gives an annual inventory of 1054.0 tCO_2e with a 95 percent interval of [1017.3, 1090.9] under Scenario A and 1053.4 [1017.0, 1089.5] under Scenario B, against a ground truth of 1050.7. Relative uncertainty improves only from 3.49 to 3.44 percent. The variance decomposition explains why: emission factors account for 92.5 percent of the variance in Scenario A and 96.7 percent in Scenario B. Better activity data cannot improve an inventory whose error budget is dominated by a published factor.

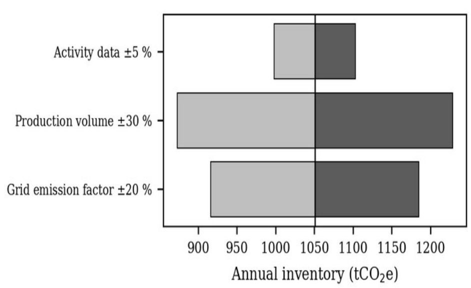

Fig. 8. One-at-a-time sensitivity of the annual inventory. The grid emission factor and production volume dominate; activity-data precision is comparatively immaterial.

Varying the grid emission factor by ±20 percent moves the inventory between 916.3 and 1185.1 tCO_2e and the emission reduction between 219.2 and 319.8 tCO_2e. Varying production volume by ±30 percent moves the inventory between 872.9 and 1228.5 tCO_2e while carbon intensity moves inversely, from 1041.7 down to 789.4 kg per 1000 parts. This inverse relation matters for disclosure: an intensity target can be met by volume growth alone, with absolute emissions rising. Spearman correlations on monthly aggregates are 0.881 between total emissions and site electricity, 0.811 against output, 0.608 against ambient temperature and −0.343 against photovoltaic yield.

TABLE IV. HYPOTHESIS TESTS AND HEADLINE OUTCOMES

H	Metric, Scenario A → Scenario B	Test	Outcome
H1	Metered machine-month cells, 0/144 → 144/144	Wilcoxon, n = 12	$p = 4.9\times10^{-4}$
H2	Median latency, 90.6 d → 8.0 min	Wilcoxon, n = 12	$p = 4.9\times10^{-4}$
H3	Machine attribution error, 11.29 → 0.33 %	Wilcoxon, n = 12	$p = 4.9\times10^{-4}$, r = 1.00
H4	Detection F1, 0.121 → 0.340; 0.540 supervised	Held-out period	Supported
H5	Traceability index τ, 0.200 → 0.984	Deterministic	Supported
–	Annual facility error, 0.31 → 0.25 %	–	No material gain
–	Inventory uncertainty, 3.49 → 3.44 %	Monte-Carlo	No material gain
–	Day-ahead carbon intensity, best R^2 = 0.005	Held-out period	Not predictable

VIII. DISCUSSION

Taken together the results support a narrower and more defensible claim than the one usually made for IIoT carbon monitoring. Continuous measurement transforms granularity, timeliness and evidentiary quality: attribution error falls by a factor of thirty-four, latency by four orders of magnitude, and the reporting matrix moves from wholly derived to almost wholly metered. It does not transform the accuracy of the headline annual inventory, because that figure was already anchored to a fiscal meter and its uncertainty is dominated by an emission factor the facility does not control.

This has a practical consequence. A facility whose objective is a defensible annual Scope 1 and Scope 2 figure gains little from machine-level instrumentation. A facility whose objective is to locate abatement, to substantiate a reduction claim, or to support product-level intensity allocation gains a great deal, because all three require resolution beneath the invoice. The credit result quantifies this: the same physical reduction is worth 21.8 tCO_2e more when the baseline can be fitted daily rather than monthly, and that difference arises purely from statistical resolution.

The negative results are equally informative. Off-the-shelf anomaly detection on raw industrial telemetry performs close to chance because industrial faults are contextual, not global outliers; a machine drawing 40 kW is normal at noon and anomalous at midnight, and no unconditional density model can express that. Day-ahead carbon-intensity forecasting fails because the numerator and denominator co-move. And an electrical-only sensing layer is structurally blind to combustion-side faults, regardless of algorithm.

IX. ESG REPORTING AND ASSURANCE IMPLICATIONS

The framework produces artefacts that map onto existing requirements without displacing them. Scope 1 and Scope 2 quantities follow the GHG Protocol operational boundaries [1]; intensity metrics satisfy GRI 305 [3] and the cross-industry metrics of IFRS S2 [2]; the tier and τ outputs correspond to the data-quality and uncertainty documentation expected under ISO 14064-1 [11]; and machine- and product-level allocation supports ISO 14067 product carbon footprints [18] and the installation-level data required for the EU carbon border mechanism [19]. For Indian reporters the monthly energy and intensity series map directly to BRSR Core attributes [4].

It must be stated plainly what the system does not do. It provides digital evidence with documented provenance; it does not provide assurance, verification or certification. A hash-chained log proves that stored records were not silently altered; it does not prove that the sensor was correctly installed, correctly scaled or correctly calibrated, and a compromised node writes false data into an unbroken chain. Independent calibration, sampling audit and third-party verification against ISO 14064-3 [12] remain necessary and unchanged. Likewise, the creditable quantity of Eq. 9 is an estimate of a potential quantity from simulated data. Additionality, baseline validity, leakage and permanence are programme-level determinations outside the scope of the measurement system, and nothing here establishes eligibility under any crediting programme.

X. LIMITATIONS AND FUTURE WORK

The evaluation is entirely simulated. No hardware was deployed and no measured data were used, so absolute error magnitudes reflect the assumed error model rather than observed instrument behaviour; the qualitative orderings are more robust than the numbers. The plant model is a single discrete-manufacturing archetype and does not represent process industries with continuous furnaces, chemical process emissions or significant fugitive sources. Scope 3 is out of scope entirely, and for most manufacturers it is the largest category. The tier weights, the deduction parameters κ and λ, and the reconstruction gap threshold are proposed values chosen for interpretability, not derived from empirical validation or any standard; different weightings would shift τ and δ without changing the structure of the argument. Emission factors are treated as time-invariant, whereas a marginal or hourly grid factor would interact strongly with the photovoltaic and load-shifting results. The fault labels used to evaluate detection are the injected ground truth, an advantage no real deployment enjoys.

Four extensions follow. First, a hardware pilot on an ESP32 and Raspberry Pi stack against reference-class meters, to replace the assumed error model with a measured one. Second, hourly marginal grid emission factors, under which load shifting rather than load reduction becomes the dominant lever. Third, extension of the tier and τ machinery to Scope 3 categories where primary data are structurally unavailable. Fourth, an empirical calibration of the deduction rule against verifier decisions, so that δ reflects observed assurance practice rather than a proposed functional form.

XI. CONCLUSION

An IIoT architecture for industrial ESG carbon reporting was developed and evaluated against conventional periodic reporting on a physics-based one-year simulation of a twelve-machine plant. Continuous 15-minute measurement reduced machine-level attribution error from 11.29 to 0.33 percent, reduced reporting latency from a median of 90.6 days to 8.0 minutes, populated the full machine-by-month reporting matrix with metered evidence, and raised the proposed traceability index from 0.200 to 0.984 at attribution level, all significant under paired non-parametric tests. Equally, three limits were quantified: the annual facility inventory improved by less than a tenth of a percentage point, total inventory uncertainty improved from 3.49 to 3.44 percent because emission factors carry over 92 percent of the variance, and day-ahead carbon intensity proved unpredictable by every model tested. The measurable value of granularity was located instead in carbon-credit quantification, where a daily rather than monthly adjusted baseline cut the conservative deduction from 14.33 to 6.24 percent and raised the creditable quantity by 21.8 tCO_2e. All results are simulated and all reproduction code, seeds and result logs are released.

ACKNOWLEDGMENT

The authors thank the Research and Development Division of Tekstein Solutions Private Limited for supporting this work.

REFERENCES

[1] World Resources Institute and WBCSD, The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard, rev. ed. Washington, DC: WRI/WBCSD, 2004.

[2] IFRS Foundation, IFRS S2 Climate-related Disclosures. London: International Sustainability Standards Board, 2023.

[3] Global Reporting Initiative, GRI 305: Emissions 2016. Amsterdam: GRI, 2016.

[4] Securities and Exchange Board of India, Business Responsibility and Sustainability Reporting Core — Framework for assurance and ESG disclosures for value chain, circular SEBI/HO/CFD/CFD-SEC-2/P/CIR/2023/122, 2023.

[5] Y. Li, X. Yang, E. Du, Y. Liu, S. Zhang, C. Yang, N. Zhang, and C. Liu, "A review on carbon emission accounting approaches for the electricity power industry," Applied Energy, vol. 359, art. 122681, Apr. 2024, doi: 10.1016/j.apenergy.2024.122681.

[6] L. J. McDonald, J. L. Hernandez Galvan, C. Emelue, A. S. S. Pinto, N. Mehta, T. Ibn-Mohammed, T. Fender, J. Radcliffe, A. Choudhary, and M. C. McManus, "Towards a unified carbon accounting landscape," Philosophical Transactions of the Royal Society A, vol. 382, no. 2282, 2024.

[7] M. S. F. Christensen, "Leveraging the Industrial Internet of Things (IIoT) for real-time CO2 monitoring, measurement and visualization: technologies, applications and future directions," in Global Internet of Things and Edge Computing Summit (GIECS 2024), Communications in Computer and Information Science, vol. 2328. Cham: Springer, 2025, doi: 10.1007/978-3-031-78572-6_3.

[8] B. Peng, Y. Li, C. Yang, H. Feng, and X. Gong, "Augmented pre-training-based carbon emission accounting method using electricity data under small-sample condition," Frontiers in Energy Research, vol. 12, art. 1505098, Nov. 2024, doi: 10.3389/fenrg.2024.1505098.

[9] M. Thornbush and A. Govind, "Monitoring, reporting, and verification (MRV) protocols used in carbon trading applied to dryland nations in the Global South for climate change mitigation," Sustainability, vol. 17, no. 24, art. 11001, Dec. 2025, doi: 10.3390/su172411001.

[10] Intergovernmental Panel on Climate Change, 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 2: Energy. Hayama, Japan: IGES, 2006.

[11] International Organization for Standardization, ISO 14064-1:2018 — Greenhouse gases — Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals. Geneva: ISO, 2018.

[12] International Organization for Standardization, ISO 14064-3:2019 — Greenhouse gases — Part 3: Specification with guidance for the verification and validation of greenhouse gas statements. Geneva: ISO, 2019.

[13] W. Jiang and H. Li, "State-of-the-art carbon metering: continuous emission monitoring systems for industrial applications," Heliyon, 2025.

[14] Central Electricity Authority, CO2 Baseline Database for the Indian Power Sector. New Delhi: Government of India, Ministry of Power, latest ed.

[15] Joint Committee for Guides in Metrology, JCGM 100:2008 — Evaluation of measurement data: Guide to the expression of uncertainty in measurement. Sèvres: BIPM, 2008.

[16] UNFCCC Clean Development Mechanism, AMS-II.D.: Energy efficiency and fuel switching measures for industrial facilities, small-scale methodology. Bonn: UNFCCC, latest version.

[17] International Electrotechnical Commission, IEC 62053-22 — Electricity metering equipment — Particular requirements — Part 22: Static meters for AC active energy (classes 0,1S, 0,2S and 0,5S). Geneva: IEC, 2020.

[18] International Organization for Standardization, ISO 14067:2018 — Greenhouse gases — Carbon footprint of products — Requirements and guidelines for quantification. Geneva: ISO, 2018.

[19] European Parliament and Council, Regulation (EU) 2023/956 establishing a carbon border adjustment mechanism, Official Journal of the European Union, L 130, 2023.

[20] International Organization for Standardization and IEC, ISO/IEC 20922:2016 — Message Queuing Telemetry Transport (MQTT) v3.1.1. Geneva: ISO/IEC, 2016.

[21] International Organization for Standardization, ISO 50001:2018 — Energy management systems — Requirements with guidance for use. Geneva: ISO, 2018.

[22] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining (ICDM), Pisa, Italy, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.