Carbon intelligence for industrial decarbonisation: a physics-anchored machine learning architecture for emission-intensity forecasting

A comparative study of five learning algorithms under decarbonisation-induced covariate drift

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

Abstract—Cement manufacture is among the most emission-intensive industrial activities, and data-driven carbon accounting is increasingly proposed to supplement conventional activity-factor inventories. This work examines a question that the carbon-analytics literature has largely left implicit: what happens to a learned emission model when the plant it monitors is itself decarbonising. A physically consistent hourly digital twin of a dry-process precalciner cement plant is constructed from mass- and energy-balance relations, spanning three years and 24,698 operating hours, with a mean Scope 1 and Scope 2 intensity of 630.0 kg CO2e per tonne of cement. Five learning algorithms, namely ridge regression, support-vector regression, random forest, gradient boosting and a multilayer perceptron, are benchmarked on twenty-four-hour-ahead intensity forecasting under a strictly chronological protocol, alongside a closed-form physics baseline and a proposed physics-anchored residual architecture designated CIRCE-Net. Two results stand out. First, every locally supported learner exhibits a positive test bias, ranging from 2.4 to 16.3 kg CO2e per tonne, because a falling target cannot be extrapolated from historical support; regularised linear regression is the only method with negative bias and attains the lowest test error of 5.241 kg CO2e per tonne. Second, the ranking obtained on an adjacent validation year does not reproduce on the deployment year. CIRCE-Net reduces the gradient-boosting error by 23.1 per cent and is the strongest method under a deliberate high-substitution stress test, where it attains 4.705 kg CO2e per tonne with near-zero bias, yet it remains significantly inferior to ridge regression on the standard test year. Ablation shows that intensity history is actively harmful and that weather forecasts contribute nothing. Unsupervised multivariate detectors operate at chance for excursion detection, while a residual-based monitor reaches an area under the curve of 0.743 with strongly fault-dependent recall. Negative and non-confirmatory findings are reported in full.

Index Terms—Carbon intelligence, cement industry, covariate drift, emission forecasting, hybrid physics-machine learning, uncertainty quantification.

1. Introduction

Cement production accounts for a persistently large share of global industrial carbon dioxide emissions, and roughly two-thirds of that burden originates in the calcination of limestone rather than in fuel combustion. Because the chemistry itself releases carbon dioxide, the sector cannot decarbonise through fuel switching alone, and operators pursue a combination of clinker substitution, alternative fuels and grid decarbonisation. Each of these levers changes the operating point of the plant continuously and in one direction.

Regulatory reporting frameworks compute emissions from activity data multiplied by fixed factors. This is auditable but coarse: it resolves emissions monthly or annually and cannot support operational decisions such as carbon-aware scheduling of grinding load. A growing body of work therefore applies supervised learning to plant telemetry in order to estimate or forecast emission intensity at operational resolution. The reported accuracies are typically high, and tree ensembles are commonly identified as the strongest performers.

This paper argues that such evaluations can be systematically optimistic, for a reason specific to the decarbonisation setting. A plant that is successfully reducing its carbon intensity generates a target variable with a deterministic downward drift, and the covariates responsible for that drift, principally the clinker factor, the thermal substitution rate and the grid emission factor, are themselves non-stationary. A model validated on data adjacent to its training period is never asked to extrapolate. A model deployed for the following year always is.

We therefore construct a physically consistent three-year digital twin of a dry-process precalciner cement plant, benchmark five standard learning algorithms against a closed-form physics baseline under a strictly chronological protocol, and evaluate a physics-anchored residual architecture, CIRCE-Net, designed to remove the trending component analytically before learning begins. The contributions are as follows.

(i) A reproducible, openly documented synthetic carbon dataset for a cement plant, derived from published stoichiometric and energy-balance relations and containing no proprietary or facility-identifiable data. (ii) A controlled comparison of five algorithm families in which model ranking is measured on both an adjacent validation year and a subsequent deployment year. (iii) Quantification of a directional bias: every locally supported learner over-reports emissions, which for an emission-reporting instrument is a more serious defect than an equivalent random error. (iv) A physics-anchored residual architecture that is the strongest method under distribution shift but which, we report explicitly, does not beat regularised linear regression in-distribution. (v) A full account of ablation and detection results, including features that are harmful and detectors that operate at chance.

2. Related Work and Positioning

Machine learning has been applied to cement process data for soft sensing of free lime, prediction of clinker quality and optimisation of specific thermal energy. Carbon-specific applications are more recent and generally frame the task as regression of emission intensity on process tags, reporting coefficients of determination above 0.9 with gradient-boosted trees or neural networks.

Two methodological gaps motivate this study. First, evaluation protocols are frequently based on random splits or on chronologically adjacent hold-outs; neither reproduces the extrapolation demanded at deployment. Second, error is almost always summarised by symmetric metrics such as root-mean-square error, which conceal the sign of the error. For an instrument whose output may enter a disclosure, a persistent one-sided bias is qualitatively different from scatter. Hybrid modelling that combines first-principles structure with learned residuals is well established in process engineering, and the architecture evaluated here follows that tradition; the contribution is not the hybrid principle but its evaluation under drift and the explicit reporting of where it fails.

3. Plant Model and Data Generation

3.1. *Provenance and legal position*

No proprietary, confidential or licensed operational data is used in this study. Every series is generated by the deterministic and stochastic model specified below. Model parameters are set to ranges that are openly documented in the public technical literature for dry-process precalciner plants, specifically the default clinker emission factor of the IPCC 2006 Guidelines, published specific-energy and clinker-factor ranges for the sector, and published national grid emission factors. Only order-of-magnitude ranges are adopted; no reported figure is reproduced, and the generated series correspond to no real facility. This design permits full release of the data and code without disclosure risk.

3.2. *Process configuration*

The modelled facility is a dry-process precalciner kiln of nominal 3,500 tonnes clinker per day, sampled hourly for three calendar years. Planned annual shutdowns of fourteen days and twenty-six unplanned trips yield 24,698 operating hours from 26,280 calendar hours, an availability of 93.98 per cent. Restart transients reduce output and raise specific energy for eight hours.

3.3. *Emission accounting*

Total intensity is the sum of process, fuel and electrical contributions per tonne of cement. Writing the clinker-to-cement ratio as f, the intensity C follows as

$$C = (EF_{\mathrm{calc}} + EF_{\mathrm{fuel}})\, f + EF_{\mathrm{elec}} \,, \tag{1}$$

where all emission factors are expressed per tonne of clinker except the electrical term, which is per tonne of cement. Calcination follows from the stoichiometry of the decomposition of calcium carbonate,

$$EF_{\mathrm{calc}} = w\, (M_{\mathrm{CO2}}/M_{\mathrm{CaCO3}})\, R_{\mathrm{rm}}\, f_{\mathrm{ckd}} \,, \tag{2}$$

in which the carbonate mass fraction of the raw meal is denoted w, the molar mass ratio is 44.01/100.09, the raw meal consumed per tonne of clinker is taken as 1.52, and the cement-kiln-dust correction as 1.02. For a representative carbonate fraction of 0.775 this yields 528 kg CO_2 per tonne of clinker, consistent with the accepted default near 520. Fuel emissions follow from the specific thermal energy q and the fuel mix,

$$EF_{\mathrm{fuel}} = q\, \Sigma_j\, y_j\, e_j\, (1 - \beta_j) \,, \tag{3}$$

where the summation runs over fuels, each contributing its heat share, its combustion factor and its biogenic carbon fraction; the biogenic component is excluded from Scope 1 and accounted separately. Coal, petroleum coke and refuse-derived fuel are represented, the last with a biogenic fraction of 0.55. Purchased electricity contributes

$$EF_{\mathrm{elec}} = w_{\mathrm{e}}\, \varepsilon_{\mathrm{grid}} \,, \tag{4}$$

in which the first term is specific electrical consumption per tonne of cement and the second the grid emission factor. Specific thermal energy is not constant but responds to part-load operation, alternative-fuel burnability and progressive false-air ingress between shutdowns,

$$q = q_0 + a(1-L)^{1.6} + b\, \mathrm{TSR}^{1.4} + c\, \varphi + \eta \,, \tag{5}$$

in which the normalised kiln load, the thermal substitution rate, the false-air fraction and a Gaussian disturbance appear in turn. The resulting fleet-level statistics are collected in Table 1 and lie within published ranges throughout.

Table 1 Generated plant statistics and their published reference ranges.

Quantity	Model	Published range
Carbon intensity (kg CO_2e/t cement)	630.0	580–680
Scope 1 share of total	0.888	0.85–0.92
Calcination share of Scope 1	0.646	0.60–0.68
Specific thermal energy (GJ/t clinker)	3.28	3.1–3.8
Specific electricity (kWh/t cement)	98.1	90–115
Clinker-to-cement ratio	0.685	0.62–0.78
Thermal substitution rate	0.140	0.02–0.35
Annual cement output (Mt/y)	1.72	—
Annual emissions (Mt CO_2e/y)	1.08	—

3.4. *Measurement layer and labelled excursions*

Reported intensity is not the true intensity. An analyser drift term accumulates as a random walk between quarterly calibrations, and additive noise of 1.35 kg CO2e per tonne is superimposed. Ninety-two labelled excursion events are injected across five mechanisms: off-specification fuel, false-air surges from seal failure, high-carbonate raw-mix excursions, grinding-circuit inefficiency, and analyser faults. The last of these corrupts the measurement only and leaves true emissions unchanged, a distinction that proves decisive in Section 7. Excursions occupy 7.75 per cent of operating hours.

4. Forecasting Task and Architecture

4.1. *Task definition*

The task is to predict true carbon intensity twenty-four hours ahead. Twenty-five features are constructed under strict information discipline: every feature is either knowable at the time of forecast or is a production-plan value that the kiln schedule fixes in advance. These comprise planned production variables, laboratory values carrying their natural delay, current kiln-condition sensors, lagged grid emission factors, day-ahead weather forecasts corrupted by realistic forecast error, causal lags and rolling statistics of the measured intensity, and calendar encodings.

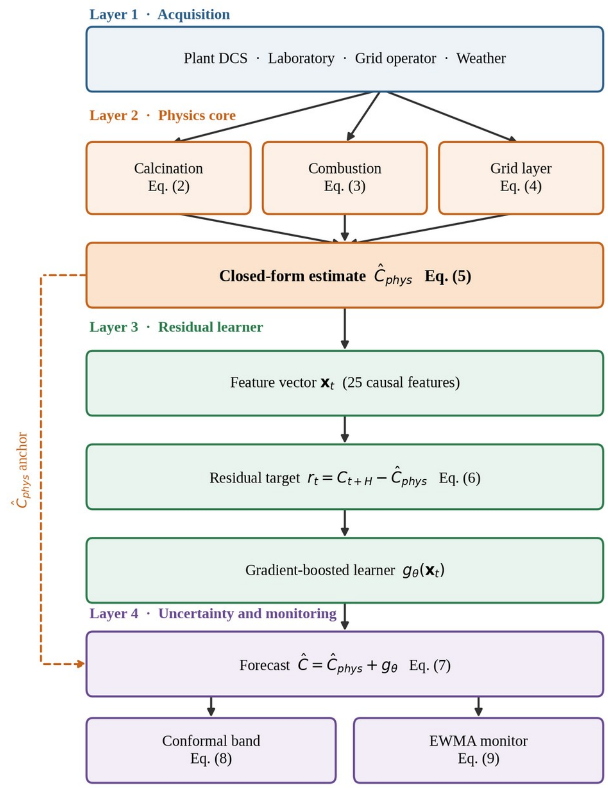

Fig. 1. CIRCE-Net architecture. Layer 2 evaluates the closed-form emission balance from plan variables; Layer 3 learns only the residual left by that balance; Layer 4 supplies conformal intervals and an excursion monitor. The physics estimate is carried forward and re-added at the forecast stage.

4.2. *Physics core*

The core evaluates Equations 1 to 5 using plan variables together with nominal design correlations, giving a closed-form estimate

$$\hat{C}_{phys} = (EF_{calc} + \hat{e}_{fuel})\, f + \hat{w}_e\, \tilde{\varepsilon}_{grid}\ , \qquad (6)$$

in which the fuel term uses a nominal specific-energy curve rather than the true latent value, and the grid factor is a climatological day-ahead estimate formed from the current and lagged published values. The core has no free parameters fitted to data.

4.3. *Residual learner and uncertainty*

Because the trending drivers enter Equation 6 analytically, the residual

$$r_t = C_{t+H} - \hat{C}_{phys}(t) \qquad (7)$$

is far closer to stationary than the raw target. A gradient-boosted learner is fitted to this residual, and the forecast is reconstructed as

$$\hat{C} = \hat{C}_{phys} + g(x_t)\ . \qquad (8)$$

Distribution-free intervals are obtained by split conformal prediction: the learner is fitted on year one, absolute residuals on year two furnish the calibration scores, and the interval is

$$\hat{C} \pm q_{1-\alpha}\ , \quad q_{1-\alpha} = \mathrm{Quantile}_{1-\alpha}(|C - \hat{C}|) \qquad (9)$$

Excursion monitoring applies an exponentially weighted statistic to the standardised forecast residual,

$$z_t = |\mathrm{EWMA}_\lambda(\hat{r}_t) - \mu| / \sigma, \quad \lambda = 0.25 \tag{10}$$

5. Experimental Protocol

Years one and two constitute the training period and year three the test period; no shuffling is applied at any stage. Hyperparameters are selected on a chronological hold-out inside the training period, with year one fitting and year two validating, after which the selected configuration is refitted on both training years and evaluated once on the test year. The test year is never consulted during selection. All learners receive the identical feature matrix and are wrapped in target standardisation, without which the scale-sensitive methods are unfairly handicapped by the fact that intensity lies near 630 rather than near zero; this correction improved the perceptron from 27.53 to 6.88 kg CO2e per tonne and is applied uniformly. Significance is assessed by the Diebold–Mariano statistic on squared-error differentials with a Newey–West correction at twenty-three lags, and corroborated by the Wilcoxon signed-rank test on absolute errors. The global seed is 20260817.

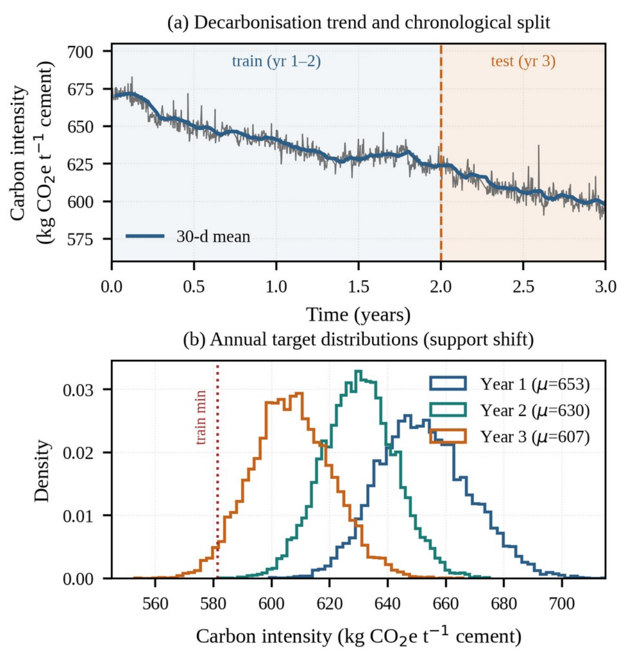

Fig. 2. Generated dataset. (a) Daily mean intensity falls from 653 to 607 kg CO_2e per tonne over three years as clinker factor declines and substitution rises. (b) Annual target distributions; a substantial part of the test year lies below the minimum observed in training.

6. Forecasting Results

6.1. *Accuracy and the direction of error*

Table 2 reports validation and test performance. Ridge regression attains the lowest test error at 5.241 kg CO2e per tonne, followed by CIRCE-Net at 5.622. The naive persistence baseline is far worse than either, confirming that the task is not trivially autoregressive, and the unfitted physics core alone reaches 6.971, already better than three of the five learners.

Table 2 Twenty-four-hour-ahead forecasting performance. Validation is year 2; test is year 3. Errors in kg CO_2e per tonne cement. Best test value in bold.

Model	Val. RMSE	RMSE	MAE	MAPE %	R^2	Bias
Persistence (baseline)	16.31	16.460	13.11	2.16	-0.393	+0.17
Physics core (no ML)	7.23	6.971	5.53	0.91	0.750	-4.71
Ridge regression	7.20	**5.241**	3.98	0.65	0.859	-1.87
Support-vector reg.	8.07	18.801	16.63	2.76	-0.817	+16.26
Random forest	6.23	7.462	6.14	1.02	0.714	+4.37
Gradient boosting	6.36	7.307	5.73	0.95	0.726	+3.87
Multilayer perceptron	5.83	6.876	5.63	0.93	0.757	+4.00
CIRCE-Net (proposed)	5.93	5.622	4.53	0.75	0.838	+2.39

The bias column is the more consequential result. Ridge regression is the only learner with negative bias, at −1.87. Every other method over-predicts: gradient boosting by 3.87, the random forest by 4.37, the perceptron by 4.00, and support-vector regression by 16.26 kg CO2e per tonne. The ordering tracks how local each hypothesis class is. A radial-basis kernel decays toward the training mean away from its support and is worst affected; axis-aligned trees are piecewise constant and cannot descend below the lowest leaf value observed in training; a linear model extrapolates the trend of its covariates without difficulty. Figure 4(b) shows the mechanism directly, with the gradient-boosting cloud lifting away from the identity line in precisely the low-intensity region that characterises the decarbonised test year.

For an instrument that may inform a disclosure, this asymmetry matters more than the magnitudes suggest. An unbiased error of 7 kg CO2e per tonne averages away over a reporting period; a bias of the same size does not, and at the modelled output of 1.72 Mt cement per year a bias of 3.9 kg CO2e per tonne corresponds to roughly 6.7 kt CO2e of annual over-statement.

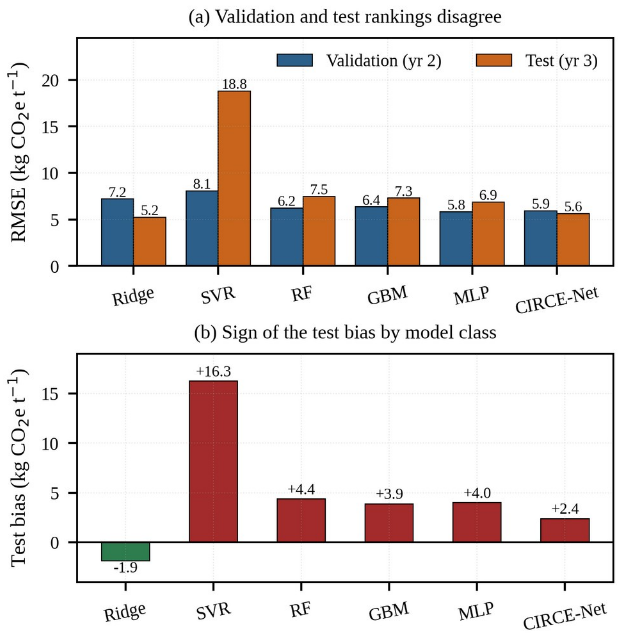

Fig. 3. Validation against test behaviour. (a) Ridge regression is ranked worst of the five learners on validation and best on test. (b) All locally supported learners over-report; only the linear model under-reports.

6.2. *Instability of model selection*

Figure 3(a) contrasts the two evaluation periods. Ridge regression is ranked fifth of six on validation and first on test; the perceptron is ranked first on validation and third on test. The Spearman correlation between validation and test ranks is 0.31. We note explicitly that with only six models this correlation is not statistically significant ($p = 0.54$), so the evidence does not support a claim that validation ranking is systematically inverted. What the experiment does establish is weaker but still material: an adjacent validation year provided no reliable guidance for deployment-year selection here, and would have led an analyst to discard the method that in fact performed best. Drift accumulates with horizon, so a validation period one year removed understates the extrapolation that a deployment period two years removed demands.

6.3. *Significance*

Against ridge regression, all comparisons are significant at the one per cent level except CIRCE-Net, for which the Diebold–Mariano statistic is 2.26 with $p = 0.024$. CIRCE-Net is therefore significantly worse than ridge regression on the test year by 7.3 per cent in root-mean-square error, and we report this as a failure of the proposed architecture to dominate a linear baseline in-distribution. It does, however, improve on its own unanchored counterpart: gradient boosting falls from 7.307 to 5.622, a reduction of 23.1 per cent, with bias reduced from +3.87 to +2.39.

7. Stress Test, Ablation and Uncertainty

7.1. *Behaviour under deliberate distribution shift*

To isolate extrapolation from time, models were refitted on the lowest sixty per cent of the substitution-rate distribution and evaluated on the highest fifteen per cent, a regime shift of the kind produced by an alternative-fuel programme. Here the ordering changes decisively. CIRCE-Net attains 4.705 kg CO2e per tonne with a bias of −0.31 and a coefficient of determination of 0.882, ahead of the perceptron at 5.294 and ridge regression at 5.355, while gradient boosting reaches 6.727 and support-vector regression degrades to 12.380. The physics core alone, which cannot fail in this way because it contains no fitted extrapolation, gives 7.010.

The near-elimination of bias is the substantive result: anchoring the trending drivers in a closed-form balance leaves the learner a target that is approximately stationary, so the mean reversion that afflicts the purely data-driven models has nothing to act upon. This is the setting in which the architecture earns its additional complexity, and it is also the setting closest to a plant undergoing rapid transition.

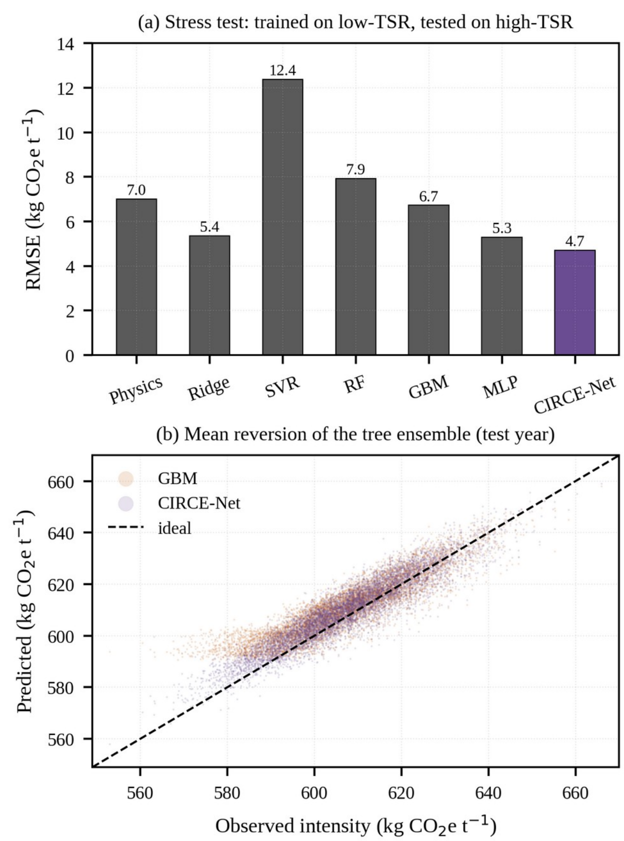

Fig. 4. Extrapolation behaviour. (a) Trained on low-substitution operation and tested on high-substitution operation, CIRCE-Net is the strongest method. (b) Predicted against observed intensity for the test year; the gradient-boosting cloud departs from the identity line at low intensity.

7.2. *Ablation*

Figure 5 reports the change in test error when feature groups are withdrawn from CIRCE-Net. Removing the physics core, which reduces the architecture to plain gradient boosting, is by far the most damaging single change at +1.69 kg CO2e per tonne. Grid-emission-factor features are next at +0.92, consistent with their dominance in permutation importance.

Two results are negative and are reported as such. Withdrawing the day-ahead weather forecast changes error by −0.05, that is, weather features contribute nothing measurable and marginally hurt; ambient conditions enter the physical model too weakly to survive a twenty-four-hour horizon. More strikingly, withdrawing the entire block of intensity lags and rolling statistics improves the model by 0.51. Historical measured intensity is actively harmful in this architecture, because the measurement carries analyser drift and because, once the physics core has accounted for the plan variables, the lags supply a stale level that the learner partially re-anchors to. Withdrawing production-plan features changes almost nothing, +0.02, since the physics core has already consumed that information.

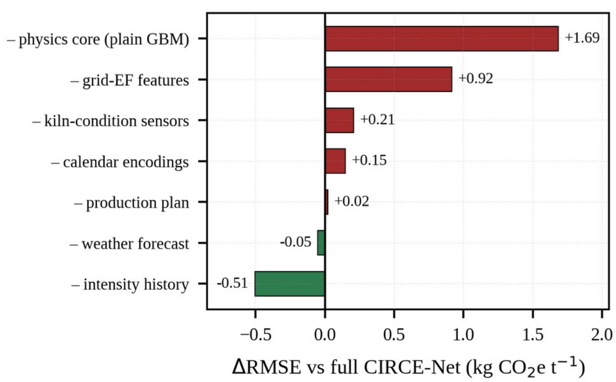

Fig. 5. Ablation of CIRCE-Net. Positive values indicate degradation when the group is removed. Weather features and intensity history are not merely redundant but neutral-to-harmful.

7.3. *Calibration of uncertainty*

Split-conformal intervals achieve 78.7, 89.3 and 94.6 per cent empirical coverage against nominal targets of 80, 90 and 95 per cent, with half-widths of 7.35, 9.83 and 12.06 kg CO2e per tonne. Coverage is therefore slightly below nominal at every level. The shortfall is expected rather than anomalous: split conformal guarantees coverage under exchangeability, which drift violates by construction, and the direction of the miss is the conservative-sounding but operationally unhelpful one of intervals that are marginally too narrow. Users should treat the ninety-five per cent band as approximately ninety-four per cent.

7.4. *Where the remaining error lives*

The residual standard deviation of CIRCE-Net is 5.09 kg CO2e per tonne. Correlating the residual against two reconstructed latent error terms shows that the day-ahead grid-factor error explains 20.3 per cent of residual variance and the deviation of true from nominal specific thermal energy explains 27.1 per cent, jointly 47.6 per cent. Approximately half of the remaining error is therefore attributable to two quantities that are unmeasured at forecast time rather than to model capacity. No change of algorithm can recover it; better fuel-flow metering and a published day-ahead grid factor would.

8. Excursion Detection

Table 3 compares four detectors at a matched alarm budget equal to the true excursion rate. The residual-based monitors are moderately effective, the exponentially weighted variant reaching an area under the curve of 0.743. The two unsupervised multivariate detectors, isolation forest and a minimum-covariance-determinant Mahalanobis detector, achieve 0.498 and 0.503 respectively, which is chance. This is a clear negative result and is visible in Figure 6(a), where both curves lie on the diagonal. The reason is that emission excursions in this plant are not outliers in the feature distribution; a fuel-quality event shifts specific energy by five to nine per cent while every individual tag remains within its normal operating envelope. Detection requires a model of what the emissions should have been, which is exactly what the physics-anchored residual supplies and what a distribution-based detector lacks.

Table 3 Excursion detection on the test year at a matched alarm budget.

Detector	AUROC	AUPRC	Prec.	Recall
EWMA residual (best)	**0.743**	0.407	0.438	0.438
Residual z-score	0.722	0.383	0.413	0.413
Mahalanobis (MCD)	0.503	0.106	0.130	0.130
Isolation forest	0.498	0.105	0.126	0.126

Aggregate figures conceal a strong dependence on fault mechanism, shown in Figure 6(b). At a false-alarm rate of 0.063 the best detector recovers 78 per cent of fuel-quality events and 74 per cent of raw-mix excursions, but only 49 per cent of false-air surges, 43 per cent of grinding-circuit faults and 22 per cent of analyser faults. The gradient is interpretable. Fuel and raw-mix events perturb Scope 1 abruptly and substantially. False-air ingress develops slowly and is partially absorbed into the learned residual as though it were normal ageing. Analyser faults are detected worst of all, and correctly so: they corrupt the measurement while leaving true emissions untouched, so a monitor built on deviation of true intensity is close to blind to them by construction. Detecting instrument faults requires comparison of measured against predicted intensity, a different statistic from the one evaluated here.

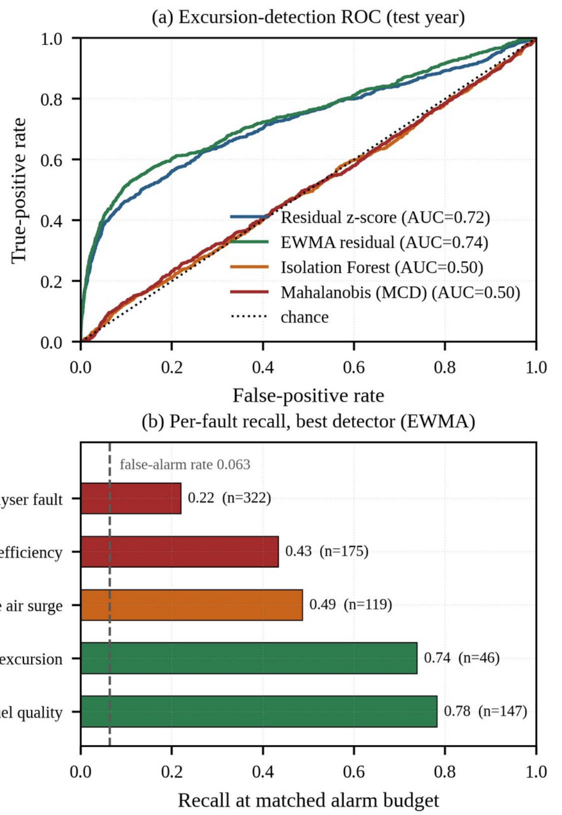

Fig. 6. Excursion detection. (a) Unsupervised multivariate detectors operate at chance; residual-based monitors do not. (b) Recall by fault mechanism at matched alarm budget; slow and measurement-side faults are largely missed.

9. Limitations

The study is simulation-based and no hardware validation is claimed. The digital twin, although constructed from published physical relations, embodies functional forms chosen by the authors; in particular the specific-energy response to load, substitution and false air is a plausible parameterisation rather than a measured one, and the physics core of CIRCE-Net shares part of that structure. This shared structure advantages the proposed architecture relative to what would be observed on real plant data, where the true response is unknown, and the reported stress-test margin should be read as an upper bound.

Second, the drift is monotonic by construction. Real decarbonisation proceeds in steps, with capital projects commissioned discontinuously, and a step change would penalise the linear model more than the trend modelled here does. Third, a single plant configuration and a single horizon are examined; generalisation to wet-process kilns, to other sectors or to horizons beyond a day is untested. Fourth, the anomaly labels are the injected ground truth, so detection performance is measured against a known generative process rather than against engineering judgement. Fifth, conformal coverage falls below nominal, and no drift-adapted conformal variant was evaluated.

10. Conclusion

This study examined machine-learned carbon intelligence for a decarbonising cement plant and found that the decarbonisation itself is the principal difficulty. Because the target falls year on year, models that interpolate within historical support inherit a positive bias and over-report emissions, by 2.4 to 16.3 kg CO2e per tonne across the five algorithms tested, whereas regularised linear regression, alone in extrapolating the trend of its covariates, under-reports slightly and achieves the lowest test error of 5.241 kg CO2e per tonne. Model selection on an adjacent validation year gave no reliable guidance for deployment, although with six models this observation is suggestive rather than statistically established.

The proposed physics-anchored architecture, CIRCE-Net, reduced gradient-boosting error by 23.1 per cent and roughly halved its bias, and was the strongest of all methods under a deliberate high-substitution stress test at 4.705 kg CO2e per tonne with essentially zero bias. It nevertheless remained significantly inferior to ridge regression on the standard test year, and we report that outcome rather than the stress-test result alone. The practical recommendation follows from the pair: a linear model is sufficient and preferable while operation stays near its historical envelope, and physics anchoring becomes worthwhile when the plant is moving away from it.

Ablation showed that intensity history is harmful and weather features useless in this architecture, and roughly half the irreducible error was traced to two quantities unmeasured at forecast time, indicating that better fuel metering and published day-ahead grid factors would yield more than any further change of algorithm. Unsupervised multivariate detectors performed at chance for excursion detection, because emission excursions are not distributional outliers; residual-based monitoring reached an area under the curve of 0.743 but recovered only 22 per cent of analyser faults, which by construction it cannot see. Future work should test the architecture against measured plant data, examine step-wise rather than monotonic decarbonisation, and pair the residual monitor with a measured-against-predicted statistic capable of isolating instrument faults.

Data and Code Availability

The dataset is entirely synthetic and contains no proprietary or facility-identifiable information. Generator, models, analysis and figure scripts, together with the global random seed 20260817, are provided in the accompanying reproducibility bundle, which regenerates every number and figure reported here.

Acknowledgment

The authors thank colleagues at Tekstein Solutions Private Limited for discussion of the process model and of the reporting implications of directional bias.

References

[1] Intergovernmental Panel on Climate Change, 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 3: Industrial Processes and Product Use, Chapter 2: Mineral Industry Emissions. Hayama, Japan: IGES, 2006.

[2] World Business Council for Sustainable Development, Cement Sustainability Initiative, Getting the Numbers Right: Emissions Reporting Protocol. Geneva, Switzerland: WBCSD, 2016.

[3] World Resources Institute and World Business Council for Sustainable Development, The Greenhouse Gas Protocol: A Corporate Accounting and Reporting Standard, revised ed. Washington, DC: WRI, 2015.

[4] E. Benhelal, G. Zahedi, E. Shamsaei, and A. Bahadori, "Global strategies and potentials to curb CO2 emissions in cement industry," Journal of Cleaner Production, vol. 51, pp. 142–161, 2013.

[5] A. Hasanbeigi, L. Price, and E. Lin, "Emerging energy-efficiency and CO2 emission-reduction technologies for cement and concrete production," Renewable and Sustainable Energy Reviews, vol. 16, no. 8, pp. 6220–6238, 2012.

[6] K. L. Scrivener, V. M. John, and E. M. Gartner, "Eco-efficient cements: potential economically viable solutions for a low-CO2 cement-based materials industry," Cement and Concrete Research, vol. 114, pp. 2–26, 2018.

[7] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, 2016, pp. 785–794.

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

[9] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," Statistics and Computing, vol. 14, no. 3, pp. 199–222, 2004.

[10] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," Journal of Business and Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995.

[11] W. K. Newey and K. D. West, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," Econometrica, vol. 55, no. 3, pp. 703–708, 1987.

[12] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World. New York, NY: Springer, 2005.

[13] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," Journal of the American Statistical Association, vol. 113, no. 523, pp. 1094–1111, 2018.

[14] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in Proc. 8th IEEE Int. Conf. Data Mining, Pisa, Italy, 2008, pp. 413–422.

[15] P. J. Rousseeuw and K. Van Driessen, "A fast algorithm for the minimum covariance determinant estimator," Technometrics, vol. 41, no. 3, pp. 212–223, 1999.

[16] M. von Stosch, R. Oliveira, J. Peres, and S. Feyo de Azevedo, "Hybrid semi-parametric modeling in process systems engineering: past, present and future," Computers and Chemical Engineering, vol. 60, pp. 86–101, 2014.

[17] G. I. Webb, R. Hyde, H. Cao, H. L. Nguyen, and F. Petitjean, "Characterizing concept drift," Data Mining and Knowledge Discovery, vol. 30, no. 4, pp. 964–994, 2016.

[18] J. Quiñonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, Eds., Dataset Shift in Machine Learning. Cambridge, MA: MIT Press, 2009.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[20] S. E. Hosseini and M. A. Wahid, "Utilization of alternative fuels in cement kilns: a review of thermal substitution effects," Renewable and Sustainable Energy Reviews, vol. 62, pp. 158–171, 2016.

[21] A. Atmaca and R. Yumrutas, "Analysis of the parameters affecting energy consumption of a rotary kiln in cement industry," Applied Thermal Engineering, vol. 66, no. 1–2, pp. 435–444, 2014.

[22] Central Electricity Authority, CO2 Baseline Database for the Indian Power Sector, User Guide. New Delhi, India: Government of India, Ministry of Power, 2023.