Physics-informed gradient-boosting prognostics of remaining useful life for PV inverters and telecom rectifiers under voltage-unbalance exposure: a simulation study

Sherbaev Javokhir Ravshan o'g'li PhD student,
Jizzakh branch of the National University of Uzbekistan named after Mirzo Ulugbek, Jizzakh, Uzbekistan javokhirsherbaev@gmail.com
ORCID 0009-0009-9351-3318 (corresponding author).
Abdumalikov Akmaljon Abduxoliq o'g'li Associate Professor,
Jizzakh branch of the National University of Uzbekistan named after Mirzo Ulugbek, Jizzakh, Uzbekistan. akmalabdumalikov6@gmail.com ORCID 0000-0001-5432-8632

ABSTRACT

Sustained voltage unbalance shortens the life of three-phase equipment at distributed sites - grid-connected photovoltaic (PV) inverters and telecommunication rectifier stations - through additional losses and accelerated thermal ageing. Because no run-to-failure dataset links long-term unbalance exposure to failures of such assets, we evaluate a hybrid pipeline in which a cumulative-damage physical model generates degradation trajectories and an XGBoost regressor learns remaining useful life (RUL) from unbalance statistics, threshold-exposure durations, thermal-load descriptors and asset context. The principal result is methodological. Fleet degradation data are clustered - 20 498 observations from 520 objects - so an observation-level split places 100 % of test rows in objects already seen in training and reports $R^2 = 0.812$. Under a fully object-disjoint protocol the model reaches $R^2 = 0.726 \pm 0.033$ over 12 generator seeds, the margin over a budget-matched Random Forest falls to 2.1 ± 2.5 RMSE percentage points, and removing the physical damage index costs 0.087 ± 0.016 of R^2 rather than the 1.03 measured under the leaky split. Conformal intervals calibrated on held-out objects attain 84.3 % coverage against a nominal 80 %, with conditional coverage of 61.5–94.5 % across RUL quintiles. A three-level balancer-recommendation layer reaches 91.6 % accuracy ($\kappa = 0.771$) against 63.9 % for a no-model threshold rule. We report the study's simulation circularity openly and treat the split-protocol comparison as the transferable finding.

Kalit so'zlar: voltage unbalance; remaining useful life; XGBoost; physics-informed machine learning; conformal prediction; data leakage; evaluation protocol; photovoltaic inverter; telecommunication power supply; predictive maintenance

Физически обоснованное прогнозирование остаточного ресурса фотоэлектрических инверторов и телекоммуникационных выпрямителей в условиях не симметрии напряжения методом градиентного бустинга: имитационное исследование

АННАТАЦИЯ

Длительная несимметрия напряжения сокращает срок службы трёхфазного оборудования распределённых объектов — сетевых фотоэлектрических (PV) инверторов и выпрямительных станций телекоммуникаций - за счёт дополнительных потерь и ускоренного теплового старения. Поскольку данных «до отказа» для таких активов не существует, оценивается гибридный конвейер: физическая модель накопленного повреждения порождает траектории деградации, а регрессор XGBoost обучается остаточному ресурсу (RUL) по признакам несимметрии и тепловой нагрузки. Главный результат — методологический: данные кластеризованы (20 498 наблюдений с 520 объектов), поэтому разбиение на уровне наблюдений даёт R^2 = 0.812; при полностью объектно-раздельном протоколе модель достигает 0.726 ± 0.033 по 12 зёрнам генератора, а удаление физического индекса повреждения стоит лишь 0.087 ± 0.016 вместо 1.03. Конформные интервалы дают покрытие 84.3 % при номинальных 80 % (61.5–94.5 % по квинтилям); трёхуровневый слой рекомендаций по симметрированию достигает точности 91.6 % (κ = 0.771) против 63.9 % у правила без модели. Симуляционная циркулярность исследования указана открыто; переносимым результатом считается вывод о протоколе оценивания

Ключевые слова: несимметрия напряжени, остаточный ресурс (RUL), XGBoost, машинное обучение с физическими признаками, конформное предсказание, утечка данных, протокол оценивания, фотоэлектрический инвертор, электропитание телекоммуникаций, предиктивное обслуживание

Kuchlanish nosimmetrikligi ta’sirida ishlaydigan fotoelektrik invertorlar va telekommunikatsiya to‘g‘rilagichlarining qolgan xizmat muddatini fizika asosidagi gradiyent busting yordamida prognozlash: simulyatsiya tadqiqoti

ANNOTATSIYA

Uzoq muddatli kuchlanish nosimmetrikligi taqsimlangan obyektlardagi uch fazali uskunalar - tarmoqqa ulangan fotoelektrik (PV) invertorlar va telekommunikatsiya to‘g‘rilagich stansiyalari - xizmat muddatini qo‘shimcha isroflar va tezlashgan issiqlik eskirishi orqali qisqartiradi. Bunday uskunalar uchun ishdan chiqishgacha kuzatilgan ma’lumotlar mavjud emasligi sababli gibrid konveyer baholanadi: jamlanuvchi shikastlanishning fizik modeli degradatsiya trayektoriyalarini hosil qiladi, XGBoost regressori esa qolgan xizmat muddatini (RUL) nosimmetriklik va issiqlik yuklamasi belgilaridan o‘rganadi. Asosiy natija metodologik: ma’lumotlar klasterlashgan (520 obyektdan 20 498 kuzatuv), shu sababli kuzatuv darajasidagi ajratish R^2 = 0.812 beradi; to‘liq obyekt-ajratilgan protokolda model 12 ta seed bo‘yicha 0.726 ± 0.033 ga erishadi, fizik shikastlanish indeksini olib tashlash esa 1.03 o‘rniga atigi 0.087 ± 0.016 ga tushadi. Konformal intervallar nominal 80 % o‘rniga 84.3 % qamrov beradi (kvintillar bo‘yicha 61.5–94.5 %); uch darajali balanslash tavsiyasi qatlami modelsiz qoidaning 63.9 % iga nisbatan 91.6 % aniqlikka (κ = 0.771) erishadi. Tadqiqotning simulyatsion aylanmaliligi ochiq bayon etiladi; asosiy ko‘chiriladigan xulosa - baholash protokoli natijasidir.

.

Keywords: kuchlanish nosimmetrikligi, qolgan xizmat muddati (RUL), XGBoost, fizika asosidagi mashinaviy o‘rganish, konformal bashorat, ma’lumotlar sizib chiqishi, baholash protokoli, fotoelektrik invertor, telekommunikatsiya elektr ta’minoti, prognozli texnik xizmat ko‘rsatish.

1. Introduction

Low-voltage three-phase networks that feed distributed assets - small grid-connected PV plants and telecommunication facilities - are rarely balanced in practice: uneven allocation of single-phase consumers, asymmetric feeder parameters and single-phase rooftop generation all displace the phase voltages from the ideal symmetric system. The standard scalar measure is the negative-sequence voltage unbalance factor, $K_{2U} = |U_2|/|U_1| \cdot 100\,\%$, obtained by Fortescue decomposition. GOST 32144-2013 requires that 10-minute averaged K_{2U} not exceed 2 % for 95 % of a week and 4 % at any time [1]; EN 50160 sets the 2 % / 95 % criterion but no universal 4 % never-exceed limit, admitting up to 3 % in some locations [2], and IEC 61000-4-30 fixes the measurement method [3].

The consequence of exceeding these limits is not an immediate fault but slow, cumulative damage, and the mechanism differs by asset class. In rotating machines, negative-sequence voltage drives currents against an impedance much lower than the positive-sequence one; NEMA MG 1 §14.36.5 states that the resulting current unbalance is of the order of six to ten times the voltage unbalance and gives a derating curve up to 5 % [19], with IEC 60034-26 as the international counterpart [20]. That mechanism applies to the air-conditioning compressor motors of a telecommunication shelter - not to the static converters themselves. A current-controlled PV inverter does not amplify voltage unbalance into proportional current unbalance; its dominant stressor is the second-harmonic ripple that negative-sequence voltage imposes on the DC link, raising capacitor core temperature through the equivalent series resistance. Wang, Davari *et al.* quantified this chain for a converter DC link and reported a B1 lifetime falling from about 12 years to about 2 years at 10 % phase unbalance for an electrolytic LC-filter configuration, a film-capacitor configuration being barely affected [21]. The temperature-to-life step follows an Arrhenius-type law whose halving interval is material-specific - near 6 °C in Montsinger's transformer work [6], codified in IEC 60216-1 [23] and, for equivalent thermal ageing, in IEEE Std C57.91 [22]. Capacitors are among the components most often implicated in converter failures, with temperature and ripple current dominant [7].

An operator of hundreds of dispersed sites therefore faces a prognostic question: given the measured unbalance exposure of each site, how much useful life remains, and where is a phase-balancing device justified? No public or, to our knowledge, private run-to-failure dataset links long-term K_{2U} exposure to failures of PV inverters or telecom rectifiers; degradation unfolds over years across heterogeneous sites, and utilities do not log component replacements against power-quality histories. The accepted response is to evaluate candidate pipelines on physics-generated degradation data first - the C-MAPSS turbofan benchmark, itself a simulator product, has anchored a decade of RUL research [8, 25] - and to proceed to field validation once the pipeline is internally consistent. That is this paper's position, and we are explicit about the limits it imposes.

The contributions are: (i) a cumulative-damage physical model of unbalance-driven ageing, fully parameterised here, producing labelled degradation trajectories for two asset classes; (ii) an analysis of the evaluation protocol itself, showing that fleet degradation data are strongly clustered by object and that an observation-level split inflates accuracy and multiplies an ablation effect by an order of magnitude; (iii) an object-disjoint benchmark against budget-matched Random Forest, ridge regression and the physical model alone, replicated over 12 seeds with object-clustered confidence intervals; (iv) conformal intervals with marginal *and* conditional coverage measured rather than assumed; and (v) a decision layer scored against explicit no-model baselines.

2. Literature review

Data-driven RUL estimation is a mature field: Si *et al.* reviewed the statistical approaches [9] and Lei *et al.* systematised the full pipeline, noting that the scarcity of run-to-failure data is its chronic bottleneck [10]. Liao and Köttig set out the taxonomy of hybrid physical/data-driven approaches this pipeline belongs to [26], and Arias Chao *et al.* demonstrated the variant used here - physics-model outputs injected as learner inputs - and released the N-CMAPSS dataset [24,25].

On the physics side, thermal ageing under unbalance is well quantified for rotating machines: de Abreu and Emanuel estimated loss of insulation life and its cost under voltage distortion and imbalance [4], and Gnaciński analysed winding temperature and life loss under unbalance combined with voltage deviation [5]. For power electronics, Wang and Blaabjerg establish temperature and ripple current as the dominant DC-link ageing stressors [7], Wang, Davari *et al.* provide the unbalance-to-lifetime chain [21], and Peyghami *et al.* survey reliability assessment for converter-dominated systems [26]. These strands supply the life models embedded in the simulator, but none yields a deployable per-site RUL estimator.

Our choice of learner follows the tabular-data evidence: XGBoost [11] is the reference gradient-boosting implementation, and two systematic comparisons found tree ensembles competitive with or better than deep networks on tabular problems [12,13]. LSTM networks estimate RUL well from long sensor sequences [14] but need trajectory volumes we will possess only after a field campaign, so we encode temporal structure through window statistics. For uncertainty we use split-conformal calibration of quantile regression [15,16].

One strand is under-represented in this literature and is central here: how the train/test split is drawn. Fleet degradation data are hierarchical - many observations per object - so a split that separates observations rather than objects lets a learner recognise objects instead of learning degradation. Splitting temporally guards against using future information but not against reusing the same assets, and we know of no unbalance-prognostics study that reports both protocols. Section 4.1 quantifies the difference on this dataset.

3. Research methodology

3.1. Cumulative-damage physical model

The target is remaining useful life in relative form, $L_{\mathrm{rel}} = \mathrm{RUL}/L_{\mathrm{nom}}$, capped at 2.5, which places assets of different ratings and nominal lives on one scale; 34 % of observations exceed 1, denoting assets that have already outlived their nominal life. The estimation problem is a regression $\mathrm{RUL} = f(\mathbf{x}) + \varepsilon$ over a feature vector $\mathbf{x}$ aggregated from monitoring data.

Labels are produced by a cumulative-damage model in the sense of Miner [17], with the damage increment given by an Arrhenius-type thermal life law rather than a fatigue-cycle law:

$$D(t) = \sum_k \frac{\Delta t_k}{L(T_{\mathrm{hot},k})}, \qquad L(T_{\mathrm{hot}}) = L_{\mathrm{nom}} \cdot 2^{(T_{\mathrm{ref}} - T_{\mathrm{hot}})/\Delta_{1/2}},$$

$$T_{\mathrm{hot}} = T_{\mathrm{amb}} + \Delta T_{\mathrm{load,max}} \cdot P^2 + k_{\mathrm{unb}} \cdot K_{2U}^2 \cdot \gamma,$$

where Δt_k is the time spent in operating regime k, P is the per-unit loading, γ is an unobserved per-object stress multiplier, and failure occurs when D reaches a per-object threshold D_{fail} (Table 1), so that nominally identical assets fail at different times. Unbalance enters through a quadratic heating term, consistent with the quadratic dependence of negative-sequence losses on K_{2U}. The implemented law is purely thermal: no Coffin–Manson cycle channel and no rainflow counting enter label generation, and the thermal-cycling feature of Section 3.2 is a descriptive count, not a damage input. Table 1 gives every parameter, without which the study is not reproducible.

Table 1. Parameters of the degradation simulator.

Parameter	Value
Nominal life L_{nom}	12 years (PV inverter), 15 years (telecom rectifier)
Unbalance heating coefficient k_{unb}	1.6 °C per %² (PV), 1.2 °C per %² (telecom)
Load-induced rise at full load $\Delta T_{\mathrm{load,max}}$	22 °C (PV), 18 °C (telecom)
Reference hotspot T_{ref}; halving interval $\Delta_{1/2}$	40 °C; 10 °C
Ambient mean; annual and diurnal amplitude	16 °C; 14 and 8 °C (PV). 18 °C; 10 and 5 °C (telecom)

Parameter	Value
Per-unit loading P	daylight profile $\times$ $U(0.60,\ 0.95)$ (PV); $U(0.45,\ 0.80)$ + 0.08 diurnal + $\mathcal{N}(0,\ 0.02^2)$ (telecom)
Base unbalance level	lognormal$(\ln 1.2,\ 0.55)$, capped at 5 %
Periodic variation	daily $U(0.10,\ 0.60)$ × base; weekly $U(0,\ 0.25)$ × base
Short-term variation	AR(1), $\varphi = 0.9$, $\sigma = U(0.08,\ 0.30)$ × base
Transient excursions	Poisson(8) per year, 4–72 h, +0.5 to +2.5 percentage points
Failure threshold D_{fail}; stress multiplier γ	lognormal$(0,\ 0.18)$; lognormal$(0,\ 0.12)$
Annual stress drift	$\prod_{y}\left(1 + \mathcal{N}(0.02,\ 0.05)\right)$
K_{2U} clipped to; horizon	0.05–8.0 %; 40 years, observations to $\min(t_{\text{fail}}$, 25 y)
Caps: D feature, L_{rel}; calendar offset	3.0, 2.5; $U(0,\ 3)$ years per object
Rated power; observation interval	3–20 kW in five steps; 60 days, at most 40 per object

Figure 1 shows the model's behaviour: under a low regime (sustained $K_{2U} \approx 0.8\,\%$) damage accumulates at essentially the nominal rate; a moderate regime ($\approx 2.2\,\%$) visibly shortens life; a severe regime ($\approx 3.8\,\%$) drives failure within a few years. Figure 2 gives the unbalance-attributable life loss for the two asset classes. The magnitudes are of the order reported for converter DC links [21]: a sustained level near the 2 % limit costs a small fraction of nominal life, and the dramatic reductions quoted in the literature belong to unbalance far above any normative limit.

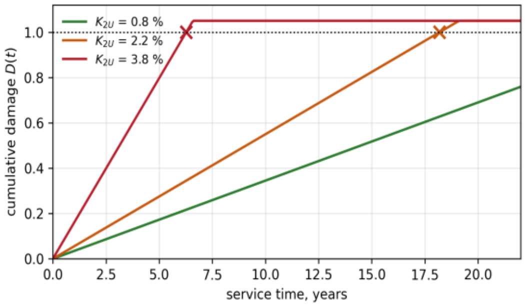

Figure 1. Cumulative-damage trajectories of the physical model for three characteristic unbalance regimes; crosses mark the failure threshold.

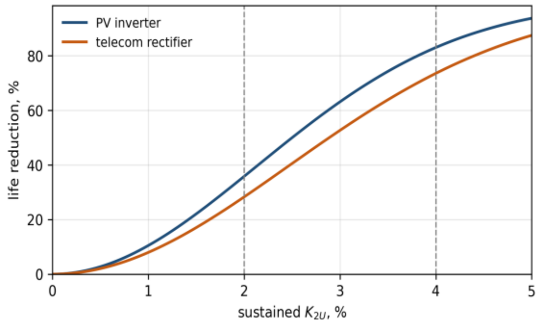

Figure 2. Unbalance-attributable life loss versus sustained K_{2U} for the two equipment classes, with the 2 % and 4 % normative limits marked.

The model plays two roles: it generates the labelled trajectories on which the learner is trained and evaluated, and at run time it supplies a physically grounded feature - the accumulated damage index, computed with γ set to unity, so the feature is a nominal estimate and the per-object stress multiplier stays hidden from the learner. The circularity this creates is analysed in Section 4.6.

3.2. Features

Features are aggregated per asset over rolling windows and fall into four groups (Table 2). The exposure features encode the dose-like character of unbalance damage - how long the asset operated above 2 % or 4 % matters more than a snapshot value - and the cumulative damage index integrates the model of Section 3.1 over the asset's history. The thermal-cycling feature counts days within the 30-day window whose temperature range exceeds 15 °C; it is a coarse range count and we do not describe it as rainflow counting.

Table 2. Input features of the RUL model.

Group	Features	Type
Unbalance statistics	k2u_mean_7d/30d; k2u_p95_7d/30d; k2u_max_30d	%
Exposure	exposure_2pct_30d; exposure_4pct_30d (dwell time above 2 % / 4 %)	hours
Physical damage	cum_damage_index (nominal Miner integral, Section 3.1)	0–3
Thermal load	temp_mean_30d; temp_max_30d; load_factor_mean_30d; thermal_cycles_30d	°C, –, count
Asset context	device_type (inverter/rectifier); service_age; rated_power	cat., years, kW

3.3. Evaluation protocol

Fleet degradation data are hierarchical: each object contributes 39.4 observations on average, and 8 of the 15 features have an intraclass correlation of at least 0.85 within an object - 3 of them are exactly constant per object. The feature vector is therefore close to an object identifier, and the protocol decision matters more than the model choice. We report three protocols: an *observation-level split* (the 70 % / 30 % calendar split of earlier versions, retained only as a leakage reference); an *object-disjoint split* (primary), 70 % of objects to training and 30 % to testing, so no test object appears in training; and an *object-disjoint and temporal split*, training on the earliest-starting 70 % of objects before the calendar median and testing on the remaining objects after it.

Baselines are budget-matched: XGBoost, the physics-free ablation and the Random Forest each receive a four-point hyper-parameter grid on the same grouped inner split, and the linear baseline is standardised ridge regression with the penalty chosen by generalised cross-validation rather than the unregularised, unscaled fit used previously. The physical model is also evaluated stand-alone. Metrics are RMSE, MAE and R^2 in relative-life units; because test rows cluster in objects, confidence intervals come from a bootstrap resampling *object*, and the protocol is replicated over 12 independent generator seeds.

3.4. Uncertainty quantification

Point estimates carry 10th–90th percentile bounds from quantile-loss variants of the model, calibrated by conformalised quantile regression [15,16]. Two details were wrong in earlier versions. The conformity quantile must be taken at the level implied by the target: for a nominal 80 % interval and n calibration points that level is $\lceil (n+1) \cdot 0.8 \rceil / n$. And the calibration set must be disjoint from the training set in the same sense as the test set — here 109 objects held out from the training objects, used neither to fit the quantile models nor to select hyper-parameters. Coverage is measured on the test set and reported conditionally as well as marginally, since a marginal figure can hide systematic failure in the segment where the tool is used.

3.5. Decision layer

The practical decision the pipeline supports is whether to install a phase-balancing device. A rule layer maps the pair (RUL estimate, observed 30-day 95th-percentile K_{2U}) to three levels: *not required*, *recommended* (30-day 95th-percentile $K_{2U} \geq 2$ % and RUL ≤ 50 %) and *necessary* ($\geq$ 4 % and RUL ≤ 30 %). The unbalance input is the observed statistic carried forward unchanged; the implementation contains no forecasting model and we make no forecasting claim. The layer is advisory. An economic payback annex promised in earlier versions of this work is withdrawn rather than reported: the implemented formula was dimensionally degenerate, the rating cancelling between saving and capital cost, and a costing study with measured prices and site tariffs is deferred to future work.

Scoring such a layer against its own reference rules is easy to do vacuously, since both arms receive the same measured K_{2U} statistic and differ only in the RUL argument. We therefore report exact per-class precision and recall, Cohen's κ, and three reference points: the majority class, a no-model threshold rule using the measured K_{2U} alone, and a constant RUL predictor.

3.6. Experimental setup

We generated 520 objects - 260 grid-connected PV inverters and 260 telecom rectifier stations - with per-object unbalance, temperature and loading scenarios drawn from the distributions of Table 1, which give a median 30-day mean K_{2U} of 1.32 % with a 5th–95th percentile range of 0.55–2.85 % and short excursions to 8.0 %, with daily and weekly periodicity. We do not claim these distributions were fitted to field measurements; they are engineering choices, and Section 4.6 treats that as a limitation. Rolling-window feature extraction yields 20 498 labelled observations; under the primary protocol 14 325 of them from 364 objects train the model and 6 173 from 156 disjoint objects test it. All code paths are deterministic under fixed seeds.

4. Analysis and results

4.1. The split protocol dominates the reported result

Under the observation-level split, all 431 test objects also occur in the training set and 100 % of test rows belong to objects the model has already seen. The split is simultaneously leaky and shifted: mean service age is 6.4 years in training against 17.8 years in test, and mean relative RUL 0.95 against 0.45, so the test population is both familiar and systematically older.

Their interaction shows in the hyper-parameter search. Tuned on a temporally split inner validation set drawn from the same objects, the search selects 600 trees at depth 6 and reports R^2 = 0.812 (RMSE 0.154); tuned on an inner split that holds out *objects*, early stopping cuts it to 60 trees and the same protocol yields 0.498 a gap of 0.31 in R^2 attributable to nothing but the inner-split rule. A deep, long-boosted model is optimal only when the objects it memorises reappear at test time.

Table 3 and Figure 3 give the model comparison under both protocols. Three conclusions from earlier versions of this work do not survive.

Table 3. Test-set metrics under the two protocols (relative-life units, primary seed). "Obs." is the observation-level split, "Object" the object-disjoint split.

Model	RMSE, obs.	R^2, obs.	RMSE, object	R^2, object
XGBoost (hybrid)	0.251	0.498	**0.259**	**0.762**
XGBoost without physics feature	0.438	-0.530	0.300	0.679
Random Forest (budget-matched)	0.249	0.504	0.267	0.746
Ridge regression	0.301	0.275	0.294	0.694
Physical model alone	0.271	0.415	0.315	0.648

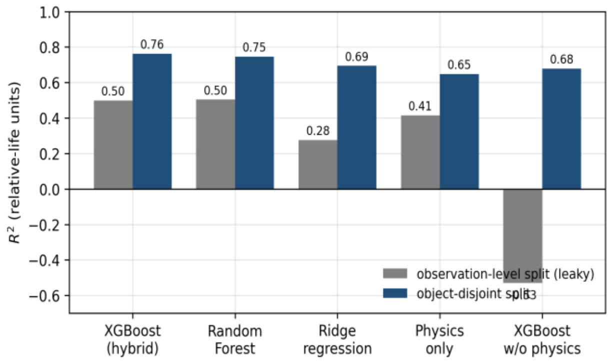

Figure 3. Explained variance by model under the observation-level and object-disjoint splits. The tuned hyper-parameters differ between protocols; both columns use the grouped inner split described in Section 3.3.

First, the claim that the hybrid model reduces RMSE by tens of percent against a Random Forest was an artefact of an untuned baseline. With the same tuning budget the reduction is 3.1 percentage points on the primary seed and 2.1 ± 2.5 points across 12 seeds. Second, the claim that linear methods fail because the problem is "essentially non-linear" does not hold: standardised ridge regression reaches R^2 = 0.694 under the object-disjoint protocol. Third, and most consequentially, the physics-feature ablation shrinks from 1.03 of R^2 under the leaky split to 0.087 ± 0.016 under the object-disjoint one. Almost the entire "collapse" reported previously measured failure to extrapolate to an older population, not the information content of the feature.

4.2. Prediction accuracy under the object-disjoint protocol

On the primary seed the hybrid model attains RMSE 0.259, MAE 0.199 and R^2 = 0.762, with an object-clustered 95 % interval of 0.72–0.80 for R^2 and 0.228–0.289 for RMSE. Across 12 independent generator seeds R^2 is 0.726 ± 0.033 (range 0.68–0.78) and RMSE 0.282 ± 0.026 (Figure 4); reporting these quantities to three significant figures from a single run, as earlier versions did, is not warranted by a seed-to-seed standard deviation of 0.033.

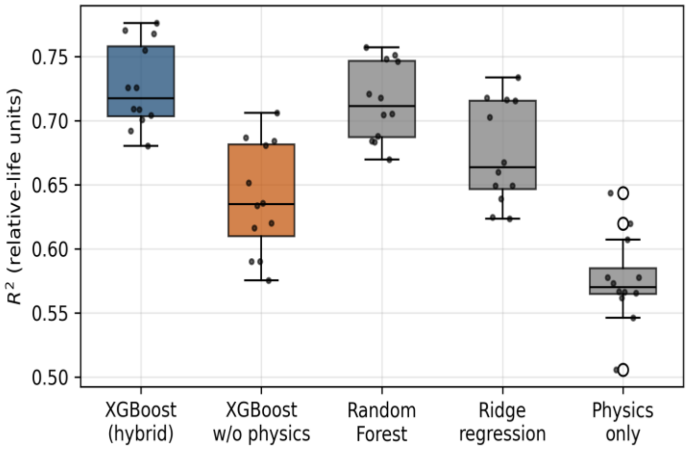

Figure 4. Explained variance across 12 independent generator seeds under the object-disjoint protocol; points are individual seeds.

An MAE of 0.199 relative-life units corresponds to about 2.4 years for a 12-year PV inverter and 3.0 years for 15-year telecom equipment - usable for annual maintenance planning, and roughly twice the error implied by the previously reported figure. Under the strictest protocol, unseen objects in a later calendar period, accuracy falls to 0.443 ± 0.082 across seeds (about 3 499 test observations from 150 objects per seed), against 0.507 for the physical model alone and 0.392 for the Random Forest. Extrapolating to older assets, not recognising unseen ones, is where the pipeline is weakest - and it is the regime a deployment enters as its fleet ages.

Earlier versions reported inference cost as about 2 μs per prediction. That is batch throughput - here 0.7 μs per sample - whereas a single-row prediction, the mode in which the platform calls the model, takes about 1.4 ms. Neither constrains the application, since features are aggregated over 10-minute intervals. The defensible computational claim is the other one: CPU-only training in minutes, a model file of a few hundred kilobytes, and in-place retraining on a modest self-hosted server.

4.3. Feature importance and physical consistency

Gain-based importance under the object-disjoint protocol, averaged over 12 seeds (Figure 5), places the cumulative damage index first at 36.1 % of total gain (± 4.4 points across seeds), followed by exposure_2pct_30d at 16.7 %, k2u_mean_30d at 13.8 % and service_age at 8.6 %. The hierarchy is consistent with a dose-like ageing mechanism, but it is a consistency check and not a discovery: the dose structure was placed in the features by us and in the labels by the simulator, so the ranking confirms that the learner uses the information it was given. Below the second rank the ordering is not stable - the asset-class indicator alone varies by 1.6 percentage points of gain across seeds — and the five K_{2U} statistics are strongly correlated, so gain is divided among near-duplicates. We therefore do not rank individual unbalance statistics against one another.

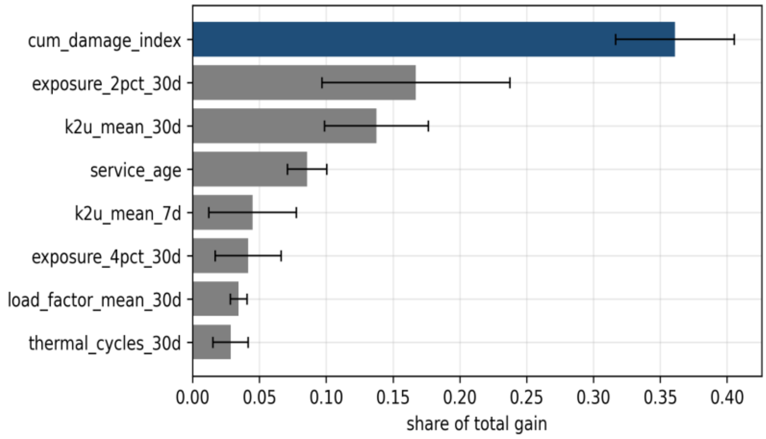

Figure 5. Gain-based feature importance of the hybrid model under the object-disjoint protocol; bars are means over 12 seeds and whiskers one standard deviation.

4.4. Calibrated prediction intervals

Raw 10th–90th percentile intervals covered 80.5 % of test targets against the nominal 80 %. Conformal calibration on 109 held-out objects (conformity quantile 0.024 in relative-life units) raised coverage to 84.3 %, with an object-clustered 95 % interval of 80.2–88.5 % and a mean width of 0.73. Two corrections to earlier versions are responsible: the conformity quantile is taken at the level implied by a nominal 80 % target rather than at 0.9, and the calibration objects are disjoint from the training objects. With the earlier procedure - a 0.9 conformity quantile calibrated on a temporal slice of the same objects - the pipeline delivered 61.9 % on the observation-level split. The shortfall was a calibration-design error, not evidence about drift.

Marginal coverage is nevertheless the wrong figure to stop at. Figure 6 gives coverage by quintile of true relative RUL: it ranges from 61.5 % to 94.5 %. The shortfall is concentrated in the longest-life quintile (L_{rel} between 1.26 and 2.50, coverage 61.5 %), where the interval is widest in absolute terms and the model is least constrained; the quintile nearest end of life, in which the interval would justify capital expenditure, is covered at 84.3 %. By asset class, coverage is 82.7 % for inverters and 85.8 % for rectifiers. The marginal figure is a fleet average; the conditional profile is the operational one.

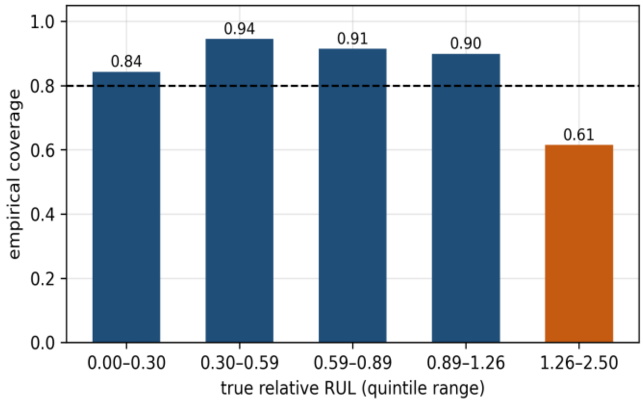

Figure 6. Empirical coverage of the calibrated 80 % prediction interval by quintile of true relative remaining useful life; the dashed line marks the nominal level.

4.5. Decision layer

Under the object-disjoint protocol the three-level recommendation matches the reference rules in 91.6 % of test observations with κ = 0.771 (Figure 7). That figure must be read against reference points. The majority class accounts for 77.2 % of observations; 43.0 % of rows fall below the 2 % threshold and return *not required* in both arms whatever the RUL error; and a no-model rule that assigns the level from the measured K_{2U} statistic alone already reaches 63.9 %. On the decisionrelevant subset - rows above the 2 % threshold - accuracy is 85.3 % against a within-subset majority of 59.9 %.

Table 4. Decision-layer performance against reference points (object-disjoint protocol, primary seed).

Predictor of RUL	Accuracy, %	κ	Recall, *nec.*	Precision, *nec.*
XGBoost (hybrid)	91.6	0.771	0.793	0.737
Random Forest	91.4	0.767	0.842	0.702
Physics only	91.5	0.756	0.813	0.794
Constant RUL 0.25	63.9	0.370	1.000	0.438
No model (K_{2U} threshold)	63.9	0.370	1.000	0.438

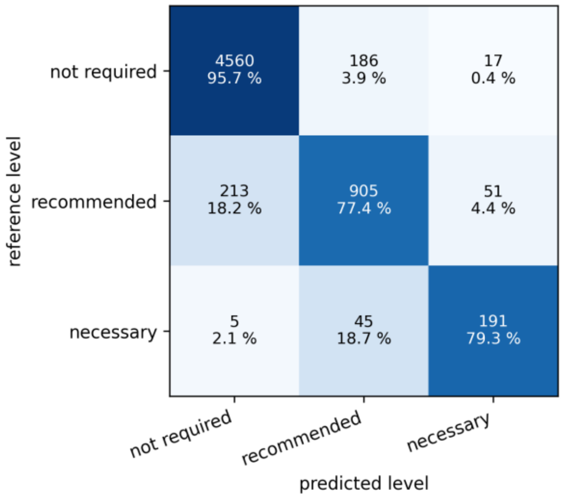

Figure 7. Confusion matrix of the three-level balancer recommendation under the object-disjoint protocol; cells give observation counts and row-normalised percentages.

Earlier versions reported that 98.2 % of safety-critical cases were flagged at least at advisory level. That metric is close to vacuous: a constant RUL predictor scores 1.000 on it against 0.979 for the trained model, because the *necessary* class is defined by a measured quantity both arms receive unchanged. On the metric with operational meaning - issuing the exact *necessary* verdict - the hybrid model recalls 0.793 of true cases against 0.813 for the physical model alone (0.656 ± 0.096 versus 0.762 ± 0.064 across seeds). The learner does not dominate the physics on the class that matters most, and we no longer claim that it does. The error anatomy is only partly favourable for advisory use: of 517 misclassified observations, 22 - 4 % - skip a level, 17 of them raising *not required* directly to *necessary* and 5 doing the reverse. The *necessary* class rests on 241 observations from 30 objects, so its percentages carry wide intervals.

4.6. Discussion and limitations

The central threat to validity is simulation circularity: the model is evaluated on data produced by the same physical assumptions that inform its strongest feature, so the headline metrics partly quantify the learner's ability to invert the simulator rather than to predict nature. The object-disjoint analysis narrows what can be claimed but does not remove this. We previously stated that scenario parameters were fitted to real monitoring distributions; that claim is withdrawn - the parameters of Table 1 are engineering choices. A field campaign is under way at a grid-connected PV installation and a telecommunication facility, where monitoring nodes log K_{2U}, its phase angle, temperature and loading continuously; maintenance records will supply genuine labels for retraining under a champion–challenger protocol.

The most informative interim experiment requires no field data: if labels are generated under one damage-law family and the physics feature is computed under another - an Eyring or inverse-power law against the Arrhenius law used here - the residual advantage of the feature measure's structure capture rather than simulator inversion. That degradation matrix is the natural next step.

Three further limitations stand: the *necessary* class is small and clustered in few objects; the simulator's damage law is single-channel and thermal, so the pipeline has never met a mechanism it does not contain; and accuracy on unseen, older assets is well below the primary figure, which is the operating point that matters most for an ageing fleet.

Overlap with related work by the same authors. The measurement node and the monitoring platform referred to here are described in two companion manuscripts in preparation, with which the simulator and the feature set of Sections 3.1–3.2 are shared. The evaluation protocol, the leakage analysis, the seed replication, the conditional-coverage analysis and the decision-layer baselines of Section 4 are specific to this paper.

5. Conclusion and recommendations

We presented a physics-informed gradient-boosting pipeline that estimates the remaining useful life of PV inverters and telecom rectifiers from voltage-unbalance monitoring data and converts the estimate into an advisory balancer recommendation. Evaluated on 520 simulated objects with an object-disjoint split, the hybrid model reaches $R^2 = 0.726 \pm 0.033$ and RMSE 0.282 ± 0.026 across 12 seeds; conformal calibration on held-out objects attains 84.3 % coverage against a nominal 80 % but varies from 61.5 % to 94.5 % across RUL quintiles; and the decision layer reaches 91.6 % agreement against 63.9 % for a no-model threshold rule. The most transferable result is not any of these figures but the demonstration that the evaluation protocol, not the model, determines what a fleet-prognostics paper reports: the same pipeline and data yield R^2 from 0.498 to 0.812, and an ablation effect differing by a factor of 12, depending on how the split is drawn. We therefore recommend the following to practitioners and researchers working with fleet monitoring data.

1. Split by object, not by observation or by date alone, and report both, since the gap measures how much of the accuracy is object recognition.
2. Budget-match the baselines before claiming a margin. An untuned Random Forest and an unscaled linear model make almost any gradient-boosting result look decisive.
3. Treat a physics-feature ablation on synthetic data as a measurement of generative leakage, not as evidence of physical information, and report it under a leakage-free split.
4. Score decision layers against explicit no-model reference points - the majority class, a threshold rule on the measured quantity and a constant predictor. A safety metric a constant predictor satisfies is not evidence.
5. Measure conditional coverage, not only marginal coverage, and log dwell time above the 2 % and 4 % limits rather than mean K_{2U}: the prognostically meaningful quantity is cumulative time above threshold, which the standards' weekly 95th-percentile rule captures only indirectly.

References

1. GOST 32144-2013. *Electric energy. Electromagnetic compatibility of technical equipment. Power quality limits in public power supply systems.* Interstate Council for Standardization, Metrology and Certification, 2013.
2. EN 50160:2010. *Voltage characteristics of electricity supplied by public electricity networks.* CENELEC, 2010.
3. IEC 61000-4-30:2015. *Electromagnetic compatibility (EMC) — Part 4-30: Testing and measurement techniques — Power quality measurement methods.* IEC, 2015.
4. J. P. G. de Abreu and A. E. Emanuel, "Induction motor thermal aging caused by voltage distortion and imbalance: Loss of useful life and its estimated cost," *IEEE Transactions on Industry Applications*, vol. 38, no. 1, pp. 12–20, 2002.
5. P. Gnaciński, "Windings temperature and loss of life of an induction machine under voltage unbalance combined with over- or undervoltages," *IEEE Transactions on Energy Conversion*, vol. 23, no. 2, pp. 363–371, 2008.
6. V. M. Montsinger, "Loading transformers by temperature," *Transactions of the American Institute of Electrical Engineers*, vol. 49, no. 2, pp. 776–790, 1930.

7. H. Wang and F. Blaabjerg, "Reliability of capacitors for DC-link applications in power electronic converters — An overview," *IEEE Transactions on Industry Applications*, vol. 50, no. 5, pp. 3569–3578, 2014.
A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. International Conference on Prognostics and Health Management (PHM)*, Denver, CO, 2008, pp. 1–9.
8. X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation — A review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, no. 1, pp. 1–14, 2011.
9. Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mechanical Systems and Signal Processing*, vol. 104, pp. 799–834, 2018.
10. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, 2016, pp. 785–794.
11. L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in *Advances in Neural Information Processing Systems 35 (NeurIPS 2022), Datasets and Benchmarks Track*, 2022.
12. R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.
13. S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *Proc. IEEE International Conference on Prognostics and Health Management (ICPHM)*, Dallas, TX, 2017, pp. 88–95.
14. Y. Romano, E. Patterson, and E. J. Candès, "Conformalized quantile regression," in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019, pp. 3538–3548.
15. V. Vovk, A. Gammerman, and G. Shafer, *Algorithmic Learning in a Random World*, 2nd ed. Cham, Switzerland: Springer, 2022.
16. M. A. Miner, "Cumulative damage in fatigue," *Journal of Applied Mechanics*, vol. 12, no. 3, pp. A159–A164, 1945.
17. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017, pp. 4765–4774.
18. ANSI/NEMA MG 1-2024. *Motors and Generators*, §14.36. National Electrical Manufacturers Association, 2024.
19. IEC 60034-26:2026. *Rotating electrical machines — Part 26: Effects of unbalanced voltages on the performance of three-phase cage induction motors*, 2nd ed. IEC, 2026.
20. H. Wang, P. Davari, H. Wang, D. Kumar, F. Zare, and F. Blaabjerg, "Lifetime estimation of DC-link capacitors in adjustable speed drives under grid voltage unbalances," *IEEE Transactions on Power Electronics*, vol. 34, no. 5, pp. 4064–4078, 2019.
21. IEEE Std C57.91-2025. *IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators*. IEEE, 2026.
22. IEC 60216-1:2025. *Electrical insulating materials — Thermal endurance properties — Part 1: Ageing procedures and evaluation of test results*, 7th ed. IEC, 2025.
23. M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physics-based and deep learning models for prognostics," *Reliability Engineering & System Safety*, vol. 217, art. no. 107961, 2022.
24. M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics," *Data*, vol. 6, no. 1, art. no. 5, 2021.
25. L. Liao and F. Köttig, "Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction," *IEEE Transactions on Reliability*, vol. 63, no. 1, pp. 191–207, 2014.
26. S. Peyghami, P. Palensky, and F. Blaabjerg, "An overview on the reliability of modern power electronic based power systems," *IEEE Open Journal of Power Electronics*, vol. 1, pp. 34–50, 2020.