

Physics-Informed Gaussian Process Regression for Battery Health Estimation Under Distribution Shift

DALITSO KUNTAMBILA ¹ AND JUN CHEN ² (Senior Member, IEEE)

¹Department of Industrial and Systems Engineering, Oakland University, Rochester, MI 48309 USA

²Department of Electrical and Computer Engineering, Oakland University, Rochester, MI 48309 USA

CORRESPONDING AUTHOR: JUN CHEN (e-mail: junchen@oakland.edu).

This work was supported by the National Science Foundation under Award #2430374.

ABSTRACT Accurate state-of-health (SOH) estimation is essential for safe, reliable battery management. Existing data-driven SOH estimation methods frequently exhibit limited operational coverage and lack calibrated predictive uncertainty under distribution shift. This study presents a physics-informed Gaussian process regression framework for SOH estimation from electrochemical impedance spectroscopy data. Equivalent circuit model parameters are extracted through a linear Kramers–Kronig validation and distribution of relaxation times deconvolution pipeline. These are combined with a Cumulative Stress Index (CSI), a scalar severity measure formed by integrating multiplicative stress factors over exposure and computable from operating history alone; it attains $R^2 = 0.938$ on holdout and 0.709 on extrapolation, without any impedance measurement; its marginal contribution alongside the impedance features is 0.027 percentage points. A training set of 64 simulated cells is selected using a vertex-anchored maximum-projection space-filling design and evaluated using Leave-One-Battery-Out cross-validation. The method achieves a holdout Mean Absolute Percentage Error (MAPE) of 0.169% ($R^2 = 0.995$) across 256 unseen simulated cells and 2.66% MAPE ($R^2 = 0.752$) on a 94-cell extrapolation set beyond the training distribution, where the error is isolated to the sub-zero temperature tier. Scalar variance recalibration attains 94.3% to 96.1% prediction interval coverage.

INDEX TERMS Battery state-of-health, electrochemical impedance spectroscopy, Gaussian process regression, distribution of relaxation times, equivalent circuit model.

I. INTRODUCTION

Accurate real-time estimation of lithium-ion battery state-of-health (SOH) is critical for safe, optimal operation. SOH quantifies residual capacity as a fraction of its initial value and serves as the primary metric for estimating remaining useful life, defining safe operational limits, and determining end-of-life (EOL) replacement schedules within battery management systems (BMS) [1], [2], [3], [4]. Electrochemical impedance spectroscopy (EIS) non-invasively probes the physicochemical state of a battery cell across a wide frequency range, resolving overlapping transport processes: electrolyte ohmic resistance, solid-electrolyte interphase (SEI) film dynamics, charge-transfer kinetics, and solid-state lithium diffusion. Linear Kramers–Kronig (LinKK) validation and distribution of

relaxation times (DRT) deconvolution establish which of these are separably resolved before any circuit model is fitted [5]. Features extracted from the spectrum predict SOH by capturing the sub-processes that govern capacity fade and impedance growth [6], [7].

Data-driven SOH estimation from EIS has employed kernel and neural network regressors, including Gaussian process regression (GPR) [8]. GPR is advantageous for battery diagnostics because it provides closed-form posterior inference, prediction intervals, feature relevance quantification, and accommodates multi-scale input features through kernel selection [9], [10]. GPR training requires $\mathcal{O}(N^3)$ computational and $\mathcal{O}(N^2)$ memory complexity with respect to the number of observations N , limiting practical training-set sizes and posing

challenges for onboard deployment [11]. Sparse and inducing-point approximations reduce this cost substantially, with a 85.7% reduction in inference time at comparable accuracy reported for a stacking-ensemble sparse GPR on fast-charging non-EIS features [12]. Existing GPR-based EIS approaches exhibit two recurring limitations: first, training sets frequently cover sparse regions of the operating space with respect to temperature, C-rate, depth of discharge (DOD), and state of charge (SOC) [12], [13]. Second, predictive uncertainty calibration is rarely validated outside the training distribution and is typically reported as interval coverage rather than as a proper scoring rule.

Covering a four-factor operating space to EOL is not attainable experimentally at reasonable cost, which motivates simulation-based fleet generation. Frameworks coupling electrochemical, thermal, and mechanical degradation pathways exist for this purpose: building on the Doyle–Fuller–Newman (DFN) formulation [14], the open-source PyBaMM library jointly captures the dominant degradation mechanisms [15], [16], [17], [18].

The Palmgren–Miner rule provides a precedent for integrating multi-stressor severity over exposure toward a normalized failure threshold [19], [20], but does not dictate how distinct stressors combine into that severity measure. Semi-empirical aging models conventionally combine Arrhenius temperature terms, exponential SOC dependencies, and power-law time progression [21]. Degradation rates derived from Eyring’s law are instead multiplicative in stress factors [22], while calendar and cycling aging interactions are non-additive [23], [24]; neither behavior is captured by the linear weighted sums common to heuristic severity indices.

This study addresses these limitations through four key contributions.

- *Stress Accumulation*: A multiplicative Stress Index (Ψ) compresses the four operating factors into a scalar severity measure whose cumulative form (CSI) reaches unity at EOL. Computable from operating history alone, it doubles as a standalone SOH predictor ($R^2 = 0.938$ holdout, 0.709 extrapolation) requiring no impedance measurement.
- *Space-Filling Anchor Selection*: Vertex-anchored maximum-projection (MaxPro) selection places training anchors at the boundary vertices of the operating space and fills the interior by a projection-aware criterion [25].
- *Three-Tier Validation*: Leave-one-battery-out (LOBO) cross-validation on 64 anchors, interpolative testing on 256 unseen cells, and per-axis extrapolation on 94 cells outside the training distribution, with interval coverage and negative log predictive density (NLPD) reported at every tier. The model is fitted to a 1,920-point stratified subsample, bounding training cost and kernel-matrix memory.
- *Kernel Structure*: A controlled comparison shows that an additive impedance/stress kernel preserves the marginal contribution of the stress covariates, which a joint Automatic Relevance Determination (ARD) kernel reverses,

TABLE 1. Simulation Variables and Operational Ranges

Factor	Interior levels (hull)	Exterior levels
C-rate (1 h^{-1})	{0.25, 0.35, 0.5, 0.7, 1.0, 1.4, 2.0}	{3.0} [†]
Temperature ($^{\circ}\text{C}$)	{0, 5, 15, 20, 30, 35, 45}	{−10}
DOD (%)	{20, 30, 40, 50, 60, 70, 80}	{90, 100}
Window position p	{0.25, 0.5, 0.75}	{0, 1}

[†]Designed but not realized: all 32 configurations at $C = 3.0 \text{ h}^{-1}$ failed to converge.

and that the association between per-cell error and distance to the nearest training anchor, significant under the joint kernel, is not detectable under the additive one.

The remainder of the paper is organized as follows. Section II covers the multi-stage data generation pipeline. Section III presents the GPR framework and feature engineering. Results and discussion are presented in Section IV. Section V concludes the paper.

II. DATA GENERATION AND IMPEDANCE CHARACTERIZATION

A. DESIGN OF EXPERIMENTS AND STRESS INDEX FORMULATION

Battery aging trajectories are simulated using the PyBaMM [15] library with a DFN pseudo-two-dimensional electrochemical model [14] and the parameter set from Chen et al. [16]. Four primary degradation sub-models from [17] are enabled alongside a lumped thermal model and a differential surface form, : (i) EC-reaction-limited SEI growth, (ii) irreversible lithium plating, (iii) Paris–Erdogan crack growth under a swelling-and-cracking particle mechanics model [18], [20], and (iv) stress-driven loss of active material (LAM). Because the degradation parameters are not included in the standard set by Chen et al., rate constants are selected within documented plausible ranges to maintain solver stability (Section V). Only SEI growth and lithium plating drive significant capacity fade in the realized fleet, with LAM and cracking channels inactive (Section IV-D). Each cell undergoes repeated discharge-rest-charge-rest cycling until its capacity drops to 70% of starting capacity or a preset cycle limit is reached. EIS is recorded every 20 aging cycles preceded by a voltage-anchored conditioning step (Section II-B), from which DRT and equivalent circuit model (ECM) features are subsequently extracted. EOL is defined at 80% SOH; 409 of the 414 converged cells reach this threshold.

The design spans four operating factors (Table 1) with physically motivated spacing: logarithmic in C-rate, uniform in inverse temperature, and linear in DOD and SOC window position. Rather than sampling SOC directly and rejecting infeasible points, the cycling window is re-parameterized as (DOD, p) with

$$\text{SOC}_{\min} = p(100 - DOD), \quad \text{SOC}_{\max} = \text{SOC}_{\min} + DOD, \quad (1)$$

so that $p \in [0, 1]$ slides the window between its lowest and highest admissible positions, ensuring every combination is

TABLE 2. Stressor Forms and Pre-Calibration Monotonicity Validation

Stressor	Form	Target	Med. ρ	Mono.	Mechanism
ψ_T	Arrhenius (cold)	time	+1.00	48/60	Li deposition [27], [28]
ψ_C	C/C_{ref}	time	-1.00	39/41	Overpotential [29]
ψ_D	DOD/DOD_{ref}	cycles	-1.00	19/24	Micro-cracking [18]
ψ_S	$SOC_{\text{avg}}/SOC_{\text{ref}}$	time	-1.00	12/12	SEI, TM dissol. [30], [31]

Mono.: matched-condition groups showing strict monotonicity, of the total tested. TM dissol.: transition metal dissolution.

feasible by construction: $SOC_{\text{avg}} = SOC_{\text{max}} - DOD/2$ follows, spanning 10 to 90%. The admissible span $100 - DOD$ contracts as DOD increases and vanishes at $DOD = 100\%$, where the window is $[0, 100]\%$ for every p and the attainable cell count is bounded by the (C, T) combinations rather than by design choice; the extrapolation partition is therefore specified across multiple axes rather than concentrated at full depth: three of the designed tiers are realized (Table 7), the fast-charge tier having failed to converge. Of 446 designed configurations, 414 converged. All 32 failures occur at $C = 3.0 \text{ h}^{-1}$, where the solver reached its minimum step size under every attempted strategy, a numerical rather than physical limitation, leaving a realized C-rate range of 0.25 h^{-1} to 2.0 h^{-1} [26].

Following the Eyring formulation [22], the four normalized stressors are combined into a Stress Index Ψ , defined as

$$\Psi = \Psi_0 \prod_{i \in \{T, C, D, S\}} \psi_i^{w_i} \quad (2)$$

Ψ_0 is a fitted scale constant ($\ln \Psi_0 = -3.855$) calibrated jointly with the elasticities. Each ψ_i is a ratio to its reference value at $C_{\text{ref}} = 1 \text{ h}^{-1}$, $T_{\text{ref}} = 25^\circ \text{C}$, $DOD_{\text{ref}} = 50\%$, $SOC_{\text{ref}} = 50\%$, so $\psi_i = 1$ at the reference condition and $\psi_i > 0$ at every attainable condition; no operating point carries zero damage, and anchoring to a reference rather than to design extremes renders Ψ comparable across designs and cell formats. Alternative power-law stressor forms such as $(C/C_{\text{max}})^\beta$ are subsumed into the estimated weights w_i , which absorb any implicit exponent. The temperature term is the cold branch of an Arrhenius factor, with $k_B = 8.617 \times 10^{-5} \text{ eV K}^{-1}$ and T in kelvin, reflecting the plating-limited regime of this cycling-dominated design. The other three are linear ratios:

$$\begin{aligned} \psi_T &= \exp \left[\frac{E_{a, \text{plat}}}{k_B} \left(\frac{1}{T} - \frac{1}{T_{\text{ref}}} \right) \right], & \psi_C &= \frac{C}{C_{\text{ref}}}, \\ \psi_D &= \frac{DOD}{DOD_{\text{ref}}}, & \psi_S &= \frac{SOC_{\text{avg}}}{SOC_{\text{ref}}}, \end{aligned} \quad (3)$$

with $E_{a, \text{plat}} = 0.35 \text{ eV}$.

The activation energy is not identifiable from lifetime calibration: it enters $\ln \Psi$ through the product $w_T E_{a, \text{plat}}$ (0.0414), so the fitted temperature elasticity scales inversely with any assumed $E_{a, \text{plat}}$. The nominal value is adopted and the calibrated model is invariant to this choice. Each functional form is validated before any parameter is estimated, using matched-condition groups of cells sharing three operating factors while the fourth varies. Table 2 reports the group-median Spearman correlation between stressor and lifetime and the number of

groups in which the relation is strictly monotone. DOD is validated in the cycle domain rather than the time domain: its median within-group correlation with cycles to EOL is -1.00 with 19 of 24 groups strictly decreasing, whereas against time to EOL it weakens to -0.50 with 9 of 24 strictly monotone. The difference follows from cycle duration, deeper cycles take proportionately longer ($\text{corr}(DOD, \text{cycle duration}) = +0.43$; median 4680 s at $DOD \leq 30\%$ against 9207 s above 70%), so a per-cycle mechanical stressor is partly offset by the reduced cycle count per hour.

Low temperature raises polarization and promotes lithium deposition during charge [27], [28], [32]. Across the simulation range -10°C to 45°C the relation is monotone in 48 of 60 matched-condition groups (Table 2), with no consistent interior minimum, so only the cold, plating-limited regime of the Arrhenius dependence is resolved. The opposing thermally activated regime, SEI growth and electrolyte decomposition, does not become rate-controlling within this range and is therefore not identified (Section IV-D). The ψ_S groups show lifetime spreads of up to $7.4\times$ at fixed C-rate, temperature, and DOD, consistent with SEI and deposit-layer growth accelerating at high SOC, where the lithiated graphite sits at low potential versus Li/Li^+ and reductive electrolyte decomposition is favored [33], [34]. The non-monotonic SOC dependence of calendar aging [30] lies outside this regime.

B. EIS ACQUISITION AND DRT-INFORMED CIRCUIT FITTING

Spectra are computed at 50 logarithmically spaced frequencies from 0.01 Hz to 10 kHz. The conditioning step charges a copy of the aged cell state at $C/3$ to the upper cut-off voltage and rests it for 30 minutes before the sweep, ensuring the aging trajectory of the original cell is unperturbed. Anchoring the measurement to a fixed voltage, rather than to the coulomb-counted top of each cell's partial-DOD window, is what couples the impedance to capacity. Under fixed-voltage anchoring the within-cell correlation between low-frequency impedance and capacity has median $|r| = 0.93$ with consistent sign across all 414 cells, whereas under coulomb-counted anchoring the electrode surface state is nearly age-invariant and the correlation sign varies between cells. The governing mechanism is the stoichiometry-window shift: at fixed terminal voltage, loss of lithium inventory displaces the electrode operating window, changing surface stoichiometry and hence exchange-current density and solid-state diffusivity monotonically with age [35], [36].

Prior to feature extraction, each spectrum is validated via the LinKK test [37], fitting a generalized Voigt circuit to the discrete empirical spectrum:

$$Z_{\text{Voigt}}(\omega) = R_0 + \sum_{m=1}^M \frac{R_m}{1 + j\omega\tau_m}. \quad (4)$$

Simulated spectra are LinKK-compliant by construction; satisfying causality, linearity, and time invariance. The test serves as a numerical verification confirming solver convergence at a well-defined state. On experimental spectra, it screens

for drift, non-linearity, and instrument artifacts. Maintaining an identical admission criterion across both datasets preserves a common feature-extraction pipeline and establishes a necessary precondition for transfer comparisons. Model order is fixed at $M = 50$, equal to number of measured frequencies, with relaxation times τ_m logarithmically spaced over $[\omega_{\max}^{-1}, \omega_{\min}^{-1}]$. Coefficients are obtained from a modulus-weighted, non-negative bounded linear least-squares solution regularized by a Tikhonov penalty $\lambda = 10^{-3}$; overfitting is controlled by the penalty and bounds rather than by order selection [37].

Validated spectra were deconvolved into a continuous DRT, $\gamma(\tau)$ [38], [39]:

$$Z(\omega) = R_\infty + \int_0^\infty \frac{\gamma(\tau)}{1 + j\omega\tau} d\tau + Z_W(\omega) \quad (5)$$

The diffusion term is a finite-length (transmissive) Warburg element:

$$Z_W(\omega) = R_d \frac{\tanh \sqrt{j\omega\tau_d}}{\sqrt{j\omega\tau_d}} \quad (6)$$

which saturates to R_d as $\omega \rightarrow 0$ and reduces to the semi-infinite form $A_w(j\omega)^{-0.5}$ at high frequency, with $A_w = R_d/\sqrt{\tau_d}$. The derived feature A_w is reported alongside the fitted parameters R_d and τ_d . The equivalent circuit below adopts the semi-infinite form: the finite-length saturation is carried by the DRT-fitted R_d and τ_d , which enter the feature set directly, and A_w links the two representations.

DRT peak coordinates initialize a modified Randles ECM [40], [41]:

$$Z_{\text{ECM}}(\omega) = j\omega L + R_s + A_w(j\omega)^{-0.5} + \frac{R_{\text{sei}}}{1 + (j\omega)^{n_{\text{sei}}} R_{\text{sei}} Q_{\text{sei}}} + \frac{R_{\text{ct}}}{1 + (j\omega)^{n_{\text{ct}}} R_{\text{ct}} Q_{\text{ct}}} \quad (7)$$

The parameter vector θ is estimated by complex nonlinear least-squares minimization of a relative-error objective [41]

$$J(\theta) = \sum_{k=1}^F \frac{|Z_{\text{meas}}(\omega_k) - Z_{\text{ECM}}(\omega_k, \theta)|^2}{|Z_{\text{meas}}(\omega_k)|^2}, \quad (8)$$

with two modifications: (i) imaginary residuals are weighted five times the real residuals, a deliberate departure from equal weighting motivated by the kinetic and diffusion arcs being resolved in the imaginary component while the real component is dominated by series resistance; (ii) a regularization term anchors the fitted branch time constants to the DRT peak positions, stabilizing the decomposition where the SEI and charge-transfer arcs overlap.

The anchor strength is set by a sweep over a fixed 40-cell subsample stratified on temperature and C-rate. Anchoring raises the mean fit residual from 2.04% to 2.11% and lowers the maximum residual on that subsample from 4.85% to 2.54%, consistent with the fleet-wide maximum of 2.55%. It is retained on that basis, since worst-case fit quality determines

clearance through the admission gates. Spectra are admitted if the ECM fit error does not exceed 5% and the polarization resistance evaluated at 1 Hz, $\text{Re } Z(\omega_1) - R_s$, agrees with the fitted component sum to within 25%; under this branch assignment the maximum fit error is 2.55% and no spectrum fails the fit-error gate.

The remaining fitting parameters are set by preliminary inspection rather than by a formal selection criterion; Tikhonov regularization level, residual weighting, and constant phase element (CPE) exponent bounds (Section V). The peak-prominence threshold, which determines the number of resolved relaxations, is selected from the sweep detailed below.

Kinetic branches are assigned by frequency window rather than by time-constant rank: the charge-transfer relaxation occupies 1 Hz to 30 Hz and the SEI relaxation 30 Hz to 300 Hz. In a 40-cell survey of 6,667 spectra, no cell crossed the 30 Hz boundary over its lifetime. Spectral content above 300 Hz, which accounts for less than 2.5% of total resistance, is merged into a combined contact-resistance and high-frequency artifact term. Rank assignment is well defined only at a fixed branch count and can arbitrarily assign different labels to the same physical process across cycles; window assignment removes that ambiguity.

The number of resolved relaxations depends on the DRT peak-prominence threshold: 46.1% of spectra resolve two peaks at a prominence of 0.10 relative to maximum, 96.8% at 0.05, and 100% at the operating value of 0.025, which lies on the stable plateau of that parameter sweep. The two CPE exponents are constrained to [0.55, 0.98] and are effectively governed by this boundary condition: n_{ct} reaches the upper bound (0.98) across all fits, consistent with the near-ideal double-layer capacitance implemented in the simulation model, where the bound prevents convergence to the limiting case $n = 1$; while n_{sei} reaches the lower bound on 71.1% of fits. These are disclosed as fit-configuration constraints, and neither exponent enters the feature set. The diffusion time constant τ_d terminates at its 0.159-s floor on 11.7% of fits, and its pinned fraction is a censoring caveat on that feature rather than grounds for exclusion.

Prior to feature extraction, 1,277 cycles across two cells are removed: late in life (beyond the 80% SOH threshold), the solver emitted zero-duration samples with terminal voltage pinned at the lower cut-off that pass all schema, monotonicity, and completeness checks but yield unbounded values under logarithmic transformation of rate features. The retained dataset comprises 414 cells and 73,422 per-cycle observations.

Fig. 1 illustrates the measurement-to-feature chain on a single cell across life: measured spectra with their ECM reconstructions at SOH 100, 90, and 80%, and the corresponding relaxation-time distributions with the SEI (30–300 Hz) and charge-transfer (1–30 Hz) assignment windows marked. The kinetic arc contracts monotonically as the cell fades, a falling impedance–SOH relationship specific to this simulation configuration (Section IV-D), and the two kinetic relaxations remain resolvably separated at every age.

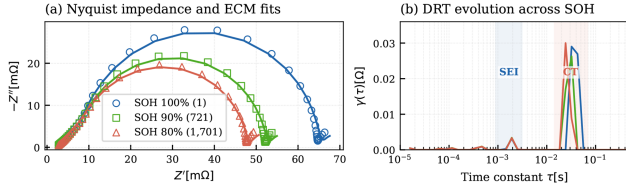


FIGURE 1. (a) Nyquist impedance (markers, LinKK-screened) and ECM reconstructions (lines) for cell C2_T15.0_D30_S65 at 100, 90, and 80% SOH (cycles 1, 721, and 1,701). (b) Corresponding DRT; shaded bands denote the SEI (30 Hz to 300 Hz) and charge-transfer (1 Hz to 30 Hz) assignment windows described in Section II-B.

III. GAUSSIAN PROCESS REGRESSION FRAMEWORK

A. FEATURE ENGINEERING AND INPUT SPACE CONSTRUCTION

The input feature space integrates localized ECM parameters with macroscale operational variables. Characteristic time constants are evaluated as:

$$\tau_{\text{sei}} = (R_{\text{sei}}Q_{\text{sei}})^{1/n_{\text{sei}}}, \quad \tau_{\text{ct}} = (R_{\text{ct}}Q_{\text{ct}})^{1/n_{\text{ct}}}. \quad (9)$$

The diffusion time constant τ_d is the finite-length Warburg parameter of Section II-B and is not CPE-derived. Natural-logarithmic transformations are applied to features whose coefficient of variation (COV) exceeds 30%; features with COV below 0.01 are excluded. Drift features $\Delta \ln X = \ln(X/X_1)$, referenced to each cell's first-cycle value X_1 , isolate degradation-driven trajectories from condition-dependent operational offsets; being per-cell log ratios, they are dimensionless and scale-independent. All features are standardized on the training distribution within each model fit.

Feature candidates are screened by a multi-criteria audit evaluating within-cell, between-cell, and pooled COV, distributional skewness, and within-cell versus pooled correlations with capacity. Within-cell and pooled correlations are evaluated separately because a feature may correlate with capacity across the fleet while carrying no aging signal within any individual cell. In such cases, the pooled statistic reflects between-cell dispersion rather than physical degradation. Fourteen features are retained:

- (i) six kinetic and diffusion levels ($R_{\text{sei}}, \tau_{\text{sei}}, R_{\text{ct}}, \tau_{\text{ct}}, R_d, \tau_d$);
- (ii) their six logarithmic drifts ($\Delta \ln R_{\text{ct}}, \Delta \ln R_{\text{sei}}, \Delta \ln \tau_{\text{ct}}, \Delta \ln \tau_{\text{sei}}, \Delta \ln R_d, \Delta \ln \tau_d$); and
- (iii) the two stress covariates (CSI and $\ln(d\text{CSI}/dN)$).

Seven candidates are excluded: the spectrum-derived and ECM-fitted series resistances, which show no physical variation; the CPE exponents n_{sei} and n_{ct} and the inductance L , which are constant or bound-determined (Section II-B); and two polarization-resistance composites, which algebraically duplicate retained features ($|\rho| \geq 0.993$). The raw kinetic levels are retained here because the identifiable branch decomposition assigns each relaxation to a consistent physical process at every cycle; under the previous, non-identifiable decomposition (Section II-B) only the drift features were usable.

Series resistance is a primary aging indicator in experimental cells [29]. As the high-frequency intercept, R_s represents the electrolyte, current-collector, and contact contributions, none of which evolve under the degradation mechanisms enabled here: its within-cell COV is below 10^{-4} and its EOL to beginning-of-life ratio is 1.000. This exclusion is therefore a property of this model configuration rather than a claim about cell physics.

B. STRESS ACCUMULATION AND CALIBRATION

Ψ of Section II-A depends only on operating conditions; cumulative exposure enters through a power-law time term:

$$\text{CSI}(t) = \Psi t^p, \quad \frac{d\text{CSI}}{dN} = \Psi p t^{p-1} \frac{dt}{dN} \quad (10)$$

where t is cumulative elapsed time and dt/dN the duration of the current cycle. Identical operating conditions yield identical Ψ ; longer exposure yields larger CSI. Within a cell, variation in $d\text{CSI}/dN$ arises from two mechanisms: the sub-linear time term t^{p-1} , which decays with exposure, and the cycle duration, which shortens as capacity fades and the cell reaches its voltage cut-offs more rapidly. The increment therefore declines over life.

Remark 1: The exponent p is not identifiable from lifetime calibration. Because the calibration condition is linear in p , the exponent vector of (2) calibrated at time exponent p , denoted $w(p)$, satisfies $w(p) = p w(1)$ exactly, so $\text{CSI}(p) = \text{CSI}(1)^p$ is a monotone power transform and both the fitted SOH–CSI relation and the leave-one-out lifetime error (13.0%) are invariant for $p \in [0.4, 1.0]$. The calibration constrains only the product pB , where B is the exponent of the SOH–CSI relation fitted during this calibration; because both quantities derive from the same capacity trajectory, this product is not an independent measurement of the fade kinetics. The value $p = 0.5$ is adopted because it yields $B = 0.957$, so that CSI is approximately proportional to capacity loss and retains the Palmgren–Miner interpretation of a consumed-life fraction reaching unity at failure [19], [42]; it is a modeling convention within the identifiable family, not a claim about fade kinetics.

As a calibration check: calibrated on the 64 training cells and evaluated on all 409 cells reaching the 80% threshold, cumulative damage at EOL has mean 0.999 and COV 11.0%, with every cell within $[0.5, 2]$. Residual correlations of EOL damage with operating factors and log lifetime are 0.280 and 0.218, respectively. Regressed on CSI alone, the index accounts for 81% of SOH variance across the fleet.

C. MODEL ARCHITECTURE AND COVARIANCE DESIGN

GPR serves as the SOH surrogate model owing to its non-parametric flexibility and closed-form predictive uncertainty [9]. The Matérn 5/2 kernel is deployed rather than the squared-exponential (RBF) kernel, as the C^∞ smoothness of the latter is a poor prior for aging trajectories containing mechanism-driven regime changes.

Impedance and stress features differ in scale and in their characteristic length scales; a single joint ARD kernel allows

one group to absorb the available amplitude. An additive structure is therefore adopted:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_1^2 k_{\text{Mat}}^{(5/2)}(\mathbf{x}_{\text{ECM}}, \mathbf{x}'_{\text{ECM}}; \ell) + \sigma_2^2 k_{\text{RQ}}(\mathbf{x}_{\text{CSI}}, \mathbf{x}'_{\text{CSI}}; \ell_s, \alpha) + \sigma_n^2 \delta_{ij} \quad (11)$$

where the first term operates on the twelve impedance features with independent length scales via ARD, and the second on the two stress covariates. Hyperparameters are estimated by maximizing the log marginal likelihood (LML) with three restarts. Length scales are bounded to $[10^{-3}, 10^2]$ and amplitudes to $[10^{-3}, 10^3]$ on standardized inputs. The GPR is fitted to a stratified subsample of 30 per-cycle observations per cell, giving 1,920 training points across the 64 anchor cells; the full 12,965 training rows are retained for LOBO evaluation and for the variance recalibration of Section IV-B.

The additive separation is consequential rather than conventional. Refitting the identical inputs, features, zero mean function, seed, and bounds with a single joint Matérn 5/2 ARD kernel over all fourteen inputs degrades holdout MAPE from 0.169% to 0.950%, LOBO MAPE from 0.242% to 1.417%, and calibrated holdout NLPD from -0.825 to $+0.808$. Kernel structure is the only difference: with a single feature block present the two configurations are the same model, and their single-block ablation configurations agree to three decimals (Table 5); they diverge only when both blocks are present. The fitted amplitudes locate the mechanism; the additive model separates them by more than five orders of magnitude (1.0×10^3 on the impedance block against 1.9×10^{-3} on the stress block), suppressing the smooth, near-deterministic stress surface within the covariance, whereas a single joint amplitude cannot express that ratio. Both configurations terminate at the same amplitude ceiling (10^3), so the comparison is not confounded by differing headroom; the additive ratio is therefore a lower bound on the block separation rather than a point estimate of it. The joint fit additionally loses ARD stability across the relevance-threshold sweep and develops a significant association between per-cell LOBO error and distance to the nearest anchor (Spearman $\rho = 0.351$, $p = 0.004$, against $\rho = 0.129$, $p = 0.31$ under the additive kernel).

A zero mean function is adopted: the GPR posterior is anchored at the standardized-target mean, and all structure is carried by the covariance. Candidate physics mean functions were screened on the training partition (Section IV-A); the best single term, $\Delta \ln \tau_{\text{ct}}$, improves the LML by 5.87 nats, clearing the 3.78-nat Bayesian information criterion (BIC) penalty, but degrades LOBO error (0.257% against 0.242%), and under the selection rule of highest marginal likelihood subject to no cross-validation degradation, no candidate is admitted. With the corrected branch decomposition the kinetic levels collectively carry the degradation trend, and an explicit trend prior is no longer required.

IV. RESULTS AND DISCUSSION

Throughout this evaluation, an *observation* denotes a single EIS spectrum reduced to its ECM parameter vector, as the

regression operates on fitted parameters rather than raw frequency points. The 414 cell trajectories (73,422 observations; Section II-B) are partitioned into three non-overlapping evaluation sets.

a) *Training Anchor Set* ($N_{\text{cells}} = 64$, 12,965 observations): Selected via 2^4 vertex anchors plus greedy MaxPro (maximum projection) infill in the coded parameter space [25]: maximin-type criteria fill only the full-dimensional space and project poorly onto subspaces, which matters when few factors are active. Model validation is conducted via LOBO-CV, decoupling hyperparameter selection from individual cell biases.

b) *Holdout Test Set* ($N_{\text{cells}} = 256$, 48,066 observations): Withheld from all feature engineering, kernel configuration, and hyperparameter optimization.

c) *Extrapolation Test Set* ($N_{\text{cells}} = 94$, 12,391 observations): Conditions at the extreme boundary of the design space outside the training distribution across three tiers: deep DOD ($\geq 90\%$), low temperature (-10°C), and SOC-window edges, testing whether physics-informed features generalize to conditions that induce qualitatively distinct degradation dynamics [43]. The fast-charge tier at $C = 3.0\text{h}^{-1}$ did not converge and extrapolation to elevated C-rate is therefore not claimed.

Point-prediction accuracy is quantified by cycle-weighted pooled MAPE, RMSE and R^2 in their standard forms; uncertainty calibration by empirical 95% prediction interval (PI) coverage, and NLPD (lower is better):

$$\text{PI}_{95} = \frac{1}{N} \sum_{i=1}^N \mathbf{1}[y_i \in [\hat{y}_i \pm 1.96\hat{\sigma}_i]] \quad (12)$$

$$\text{NLPD} = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{2} \ln(2\pi\hat{\sigma}_i^2) + \frac{(\hat{y}_i - y_i)^2}{2\hat{\sigma}_i^2} \right] \quad (13)$$

LinKK screening rejects no spectrum: the maximum relative residual across all 73,422 spectra is 0.484% (mean 0.258%); admission is governed by the equivalent-circuit gates of Section II-B, which no post-correction fit fails. The decomposition is identifiable at every cycle: the two kinetic relaxations are separated by at least 0.905 decades across all 73,422 spectra, no fit collapses the branches onto a common time constant, and no adjacent-cycle change in the kinetic drift features exceeds 0.502, under a quarter of the 2.08 in $\Delta \ln \tau$ that interchanging the branches would imply. Identifiability is the stronger criterion: a degenerate decomposition can satisfy a residual tolerance while assigning the branches arbitrarily between cycles.

A. FEATURE RELEVANCE AND MEAN FUNCTION SELECTION

Feature relevance is quantified by leave-one-feature-out (LOFO) refitting: the model is refitted without each feature in turn and the change in pooled held-out error recorded. Refitting rather than perturbing a fixed model allows the remaining features to absorb the loss, so the reported change measures

TABLE 3. Leave-One-Feature-Out Relevance: Change in Pooled MAPE When Each Feature is Removed and the Model Refitted on the Training Subsample. Positive Values Indicate the Feature is Contributing.

Feature	Δ MAPE (pp) holdout	Δ MAPE (pp) extrapolation
Diffusion time constant (τ_d)	0.0301	-0.0236
SEI time constant (τ_{sei})	0.0280	0.0201
Diffusion resistance (R_d)	0.0189	-0.0176
CT time constant (τ_{ct})	0.0157	0.0402
CT resistance (R_{ct})	0.0145	0.0175
Cumulative Stress Index (CSI)	0.0088	0.0172
SEI resistance (R_{sei})	0.0028	-0.0822
Per-cycle damage ($\ln(dCSI/dN)$)	0.0024	0.0058
Diffusion resistance drift ($\Delta \ln R_d$)	0.0011	-0.0040
SEI time-constant drift ($\Delta \ln \tau_{sei}$)	0.0009	0.0002
Diffusion time-constant drift ($\Delta \ln \tau_d$)	0.0001	0.0001
CT time-constant drift ($\Delta \ln \tau_{ct}$)	-0.0026	0.0002
CT resistance drift ($\Delta \ln R_{ct}$)	-0.0029	0.0019
SEI resistance drift ($\Delta \ln R_{sei}$)	-0.0056	-0.0021

Baselines: holdout 0.169%, extrapolation 2.660%. One refit per feature, no repeats; fixed seed, no restarts.

what a feature contributes that no other feature supplies. The procedure is deterministic: one refit per feature, no repeats.

No single feature is individually necessary (Table 3): the largest LOFO effect is 0.030 percentage points (τ_d) against a holdout baseline of 0.169%, and three drift features carry small negative values, indicating that their removal is marginally beneficial. Relevance is nonetheless ordered by block. The six ECM level features occupy the first five positions and the seventh, summing to 0.110 percentage points; the six scale-free drift features occupy the last six, summing to -0.009 percentage points; the two stress covariates fall between them at ranks six and eight. This ordering is consistent with the identifiable branch decomposition of Section II-B, in which each relaxation is assigned to a consistent physical process at every cycle, and the drifts add largely redundant information. The contrast with the block ablation of Table 5 is the substantive result: removing any one feature costs at most 0.030 percentage points, whereas removing the stress block costs 0.027 percentage points and the impedance block 1.945 percentage points. Predictive relevance is more apparent at the block level rather than at individual features, which provides empirical motivation for modeling the impedance and stress groups as separate covariance components. An additive covariance structure is adopted in Section III-C.

ARD is not used to rank features, because its per-feature amplitude allocation is not identifiable in this configuration: across near-equivalent fits the stress block receives between 0.02% and 99.93% of the total amplitude. Because ARD and LOFO quantify different aspects of relevance, their rankings are not expected to coincide. Where the two measures can be compared they agree only in part; ARD and LOFO both rank τ_d first, but ARD places the stress covariates second and third where LOFO places them sixth and eighth. The ARD-retained feature set is nonetheless stable, invariant across threshold multipliers 2.0–5.0 (nine features at every threshold: the six levels and the stress covariates plus $\Delta \ln R_{sei}$), and a single feature, $\Delta \ln \tau_d$, terminates at the length-scale upper bound

TABLE 4. Mean-Function Selection. LML Screening on the Training Partition; LOBO-CV and a Held-Out high-CSI Extrapolation Guard for the Finalists. The Guard is Evaluated Without Variance Recalibration and Ranks Candidates Only.

Mean function	k_β	Δ LML (nats)	LOBO-CV		Guard MAPE (%)
			MAPE (%)	R^2	
Zero mean (adopted)	1	—	0.242	0.9963	0.281
$\Delta \ln \tau_{ct}$	2	+5.87	0.257	0.9961	0.204
best two-term ($+\tau_{ct}$)	3	-35.37	—	—	—

TABLE 5. Feature Block Ablation Under the Zero Mean Function. The Final Row is a Sensitivity Refit Differing From the Frozen Model in Kernel Structure Alone.

Feature block	Holdout		Extrapolation	
	MAPE	R^2	MAPE	R^2
Stress only (CSI, $\ln dCSI/dN$)	2.114%	0.938	4.496%	0.709
Impedance only (12 features)	0.196%	0.993	2.677%	0.750
Both (frozen model)	0.169%	0.995	2.660%	0.752
Both, single joint kernel	0.950%	0.972	4.617%	0.580

and is pruned by ARD at all thresholds. It is reported on that basis alone.

Mean-function candidates were screened by LML on the training partition over the feature pool: eleven single-term candidates and pairwise additions to the stage-1 leader, with LOBO-CV for the finalists (Table 4). Selection follows a fixed rule: highest LML, an additional coefficient admitted only where the gain exceeds the 3.78-nat ($\frac{1}{2} \ln 1,920$, the GPR training size) BIC penalty, and no LOBO-CV degradation. Only $\Delta \ln \tau_{ct}$ clears the LML gate (Δ LML = +5.87); no pairwise addition clears the penalty. Carried to cross-validation, however, it degrades LOBO error from 0.242% to 0.257%, and the zero mean is therefore retained. On the held-out high-CSI extrapolation guard the rejected candidate is in fact slightly stronger (0.204% against 0.281%); the observed out-of-distribution cost of the zero mean on this split is 0.08 percentage points, with near-identical minimum predicted SOH (67.7 against 67.5%); the selection rests on the cross-validation criterion.

The relative contribution of each feature block is isolated in Table 5 under the frozen zero-mean configuration. The stress covariates alone attain 2.114% holdout MAPE ($R^2 = 0.938$, extrapolation $R^2 = 0.709$) from operating history alone, while the impedance features alone attain 0.196% ($R^2 = 0.993$). Given the full impedance set, the marginal contribution of the stress block is +0.027 percentage points on holdout (0.196% $\rightarrow$ 0.169%) and +0.017 under extrapolation: a high standalone R^2 with a small marginal gain indicates redundancy with the impedance features where those are available, not absence of information. That gain is contingent on the additive structure: the same two covariates added to the same twelve impedance features under a joint kernel cost 0.754 percentage points rather than gaining 0.027. Where a dense frequency sweep is unavailable, the stress block alone remains a usable predictor.

TABLE 6. Performance Summary (Cycle-Weighted Pooled)

Tier	Cells	N	MAPE	R^2	PI ₉₅		NLPD	
					Raw	Cal.	Raw	Cal.
LOBO-CV	64	12,965	0.242%	0.996	60.6%	95.0%	—	—
Holdout	256	48,066	0.169%	0.995	63.3%	94.3%	1.517	−0.825
Extrapolation	94	12,391	2.660%	0.752	76.5%	96.1%	2.060	0.807

Calibrated LOBO PI₉₅ equals the target by construction: the inflation factor is estimated on LOBO z-scores. NLPD is reported only where independent of the calibration source.

B. PREDICTIVE ACCURACY AND UNCERTAINTY CALIBRATION

Table 6 reports point accuracy and prediction-interval performance across all evaluation partitions. Per-cell error within the holdout set shows no significant association with distance to the nearest training anchor (Spearman $\rho = -0.037$, $p = 0.56$; LOBO: $\rho = 0.129$, $p = 0.31$), indicating that error is associated with crossing the domain boundary rather than with distance decay inside it.

Raw 95% prediction intervals are substantially narrower than nominal, attaining 63.3% pooled coverage on the holdout set; per cell, the median raw coverage is 62.8% with interquartile range [48.6, 75.9] across the 256 holdout cells, so the under-coverage is systematic rather than driven by a few cells. A single scalar variance-inflation factor of 2.887—the scalar matching nominal coverage of out-of-sample LOBO z-scores on the training cells only, applied unchanged to every tier—restores coverage to 95.0% (LOBO), 94.3% (holdout), and 96.1% (extrapolation), and reduces holdout NLPD from 1.517 to −0.825 (Table 6). A single global factor is an empirical adequacy finding of this fleet, not a structural guarantee under covariate shift. Negative calibrated NLPD reflects average predictive densities above unity, as expected for sub-percent residuals on a percentage-scale target. Its adequacy across partitions can be read from the scale each would require on its own residuals: 3.025 on the holdout set, 2.019 on the SOC-window edge, 3.612 on the deep-DOD tier and 2.592 on the extrapolation set as a whole. The single value of 2.887 lies within this range and holds coverage between 91.0% and 97.7% on all in-hull and moderate-shift partitions. Removing one training anchor at a time and re-estimating gives a range of 2.724 to 3.016; a bootstrap over the 64 training cells gives a mean of 2.916 with a 95% interval of [2.405, 3.547], so the factor is reasonably robust to deletion of any individual cell while uncertainty in the estimated tail quantile remains non-negligible. The sub-zero tier is the exception and requires a factor below unity (0.249), its raw intervals being already wider than nominal coverage demands. The observed pattern is consistent with two structural contributors: the homoscedastic noise assumption limits variance adaptation within the training hull, while the distance-dependent prior variance increases uncertainty far outside the training domain. A single global scalar corrects the first and does not provide uniform calibration across all regimes; input-dependent noise

TABLE 7. Extrapolation Performance by Tier: Cycle-Weighted MAPE and Raw Prediction Intervals. Per-Cell R^2 on the Cold Tier Ranges From −0.40 to −0.02; Tier-Pooled R^2 is Reported Only for the Aggregate.

Tier	Cells	N	MAPE	R^2	PI ₉₅ raw
SOC-window edge	24	5,331	0.121%	—	74.6%
Deep DOD ($\geq 90\%$)	40	4,007	0.233%	—	61.0%
Cold (-10°C)	30	3,053	10.279%	< 0	100.0%
All extrapolation	94	12,391	2.660%	0.752	76.5%

or a heteroscedastic likelihood would be required to address both, and is left to future work.

The aggregate extrapolation MAPE (2.66%) is attributable entirely to one tier (Table 7). The SOC-window-edge and deep-DOD ($\geq 90\%$) tiers perform at in-distribution accuracy 0.121% and 0.233% cycle-weighted MAPE, and the 64 non-cold extrapolation cells pool to 0.169%, identical to the holdout figure. The cold tier (-10°C) reaches 10.28% with negative per-cell R^2 throughout: this is the one extrapolation direction in which the dominant degradation mechanism changes rather than intensifies, as lithium plating becomes rate-limiting below 0°C [27], a regime absent from the training data, whose lowest temperature anchor is 0°C . Raw interval coverage on that tier is 100% for every cell, so the model reports elevated uncertainty rather than issuing confident incorrect predictions. This behavior suggests a practical use for the predictive variance beyond interval reporting: raw coverage reaches 100% on that tier and the required inflation factor falls below unity, so the variance widens sharply on entering a regime whose governing mechanism is absent from training. The increased predictive variance is consistent with the model responding to this novel regime, although that response has not been validated as a formal novelty detector.

Fig. 2 shows representative holdout SOH trajectories. Holdout cells with at least 50 observations are ordered by MAPE, and the best, median, and 90th percentile are selected. The panel is reproducible rather than curated; the 90th percentile is used instead of the maximum, which is a single pathological cell (2.27% against a fleet median of 0.087%). The best and median cells (C0.5_T5.0_D80_S95, 0.019%; C0.25_T15.0_D20_S80, 0.083%) track the trajectory to within the marker width. The 90th-percentile cell (C0.7_T35.0_D70_S77.5, 0.360%) exhibits one isolated excursion at which the posterior variance rises sharply, coinciding with an adjacent-cycle jump in the diffusion-branch features (Section II-B).

C. TRANSFER TO MEASURED CELLS

The framework was applied to 2,590 impedance spectra from twelve commercial LiCoO₂/graphite cells cycled at 25, 35, and 45°C [13], separating transferability of the EIS feature-extraction pipeline from that of the learned impedance-SOH mapping.

The feature-extraction pipeline transfers. LinKK screening rejects no spectrum (maximum relative residual 0.812%);

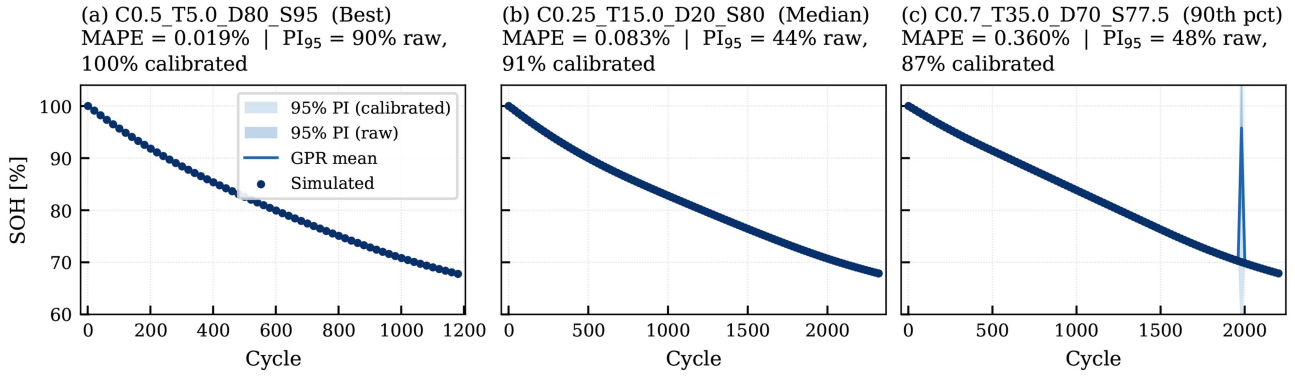


FIGURE 2. GPR posterior mean and 95% prediction intervals (raw and calibrated) against simulated SOH for 3 representative holdout cells selected by a fixed rule (best, median, 90th percentile of per-cell MAPE; $n \geq 50$). SOH is defined against each cell's first-cycle discharge capacity ($\text{SOH}_{1,2}$). Prediction intervals are narrower than the marker size on well-predicted cycles and are visible only where the posterior variance rises; raw intervals under-cover (Table 6) and the calibrated band applies the global factor 2.887.

TABLE 8. Transfer to Twelve Commercial cells [13]. The First Three Configurations Use No Measured Supervision; the Last Two are Refitted on Six Cells and Evaluated on the Remaining Six.

Configuration	MAPE (%)	R^2	PI_{95} (%)
Frozen model, all 14 features	16.44	-0.176	100.0
Impedance features only	16.45	-0.177	100.0
Scale-free drift features only	16.57	-0.197	100.0
Refit on six cells, stress only	24.66	-1.972	93.5
Refit on six cells, impedance + stress	7.73	0.694	92.7

equivalent-circuit fits have a mean error of 1.69% and a maximum of 3.85%; and both kinetic branches are detected in all spectra at prominence thresholds of 0.025 and 0.05, and in 95.5% at 0.10. Of the 2,590 spectra, 23 (0.89%) are excluded for a non-finite kinetic branch fit, leaving 2,567 evaluation samples.

The learned mapping does not transfer. Nine of fourteen features lie outside their simulated training ranges, several by two orders of magnitude. The frozen model yields 16.44% MAPE ($R^2 = -0.176$), below the experimental sample-mean predictor (Table 8). Restricting the model to scale-free drift or impedance features does not recover accuracy, indicating that the failure is not attributable to feature availability. The dominant discrepancy is a sign reversal in charge-transfer time-constant drift: $\Delta \ln \tau_{ct}$ spans $[-0.62, 0]$ across the 414 simulated cells but $[-2.03, +2.17]$ in the measured cells. Because SEI film resistance is absent from the linearized impedance and mechanical degradation channels are inactive in the simulation (Section IV-D), the stoichiometry-window shift is the primary active impedance-aging mechanism and exhibits the opposite trend at the voltage-anchored measurement state. Refitting the same architecture on six cells improves performance to 7.73% MAPE ($R^2 = 0.694$) on the remaining six, showing that the measured features contain predictive signal when the mapping is re-estimated for the target fleet. Prediction intervals achieve 100% coverage for the frozen model, with increased predictive variance outside the training domain.

This twelve-cell, single-chemistry dataset does not establish general transferability. The results indicate that the primary limitation is a mismatch between the degradation mechanisms represented in the simulator and those expressed in measured impedance. Determining which mechanisms must be added to the simulated impedance to obtain transferable SOH mappings requires a dedicated study.

D. LIMITATIONS

Four scope constraints bound the generalizability of these results. First, all training data derive from a single NMC/graphite DFN parameterization [16]; the optimized kernel hyperparameters reflect that system's kinetics. Validation across chemistries, form factors, and electrolyte formulations is a prerequisite for deployment in a physical BMS. Second, simulation fidelity constrains transfer: SEI film resistance does not enter the linearized impedance; the LAM and cracking channels are inactive (fleet maxima below 4×10^{-7} % of capacity), a consequence of the particle-cracking rate constant lying toward the low end of its documented range (Section V); and the $C = 3.0 \text{ h}^{-1}$ tier is numerically intractable under the present solver (Section II-A). At higher charging rates lithium plating is expected to become increasingly important relative to the secondary channels represented here [27]; because the coupled degradation model does not converge at $C = 3.0 \text{ h}^{-1}$, that regime is not characterized quantitatively. Extending the framework to it would require a rate-dependent C-rate stressor rather than the linear ratio calibrated here, and a plating-related impedance contribution absent from the current feature set; the additive kernel structure itself would not require modification. Because the mechanical channel is inactive, the estimated DOD elasticity ($w_D = 0.122$) and the impedance aging signature are expected to understate their active-cracking counterparts. These constraints determine the sign and magnitude of the simulated impedance-capacity-fade relationship; the solver limit further restricts the realized C-rate range to 0.25 h^{-1} to 2.0 h^{-1} . The simulated spectra are additionally free of instrumentation artifacts, inductive tails,

and stochastic baseline noise; the scale-free drift features are the components most exposed to these, being adjacent-cycle differences that amplify measurement noise simulated trajectories do not contain. Both constraints bear on the transfer behavior reported in Section IV-C. Third, the stress index accumulates without stressor-sequence memory, precluding knee-point acceleration attributed to LAM and cracking; because the mechanical channel is inactive in this fleet, no knee-point behavior is present in the training data for the index to have captured. Cumulative exposure is undifferentiated elapsed time, each cycle contributing its full duration, rest intervals included, at a single Stress Index, so cycling and calendar contributions are not separable and the C-rate elasticity absorbs both, with the rest fraction of a cycle co-varying with C-rate and drifting as capacity fades; its temperature term is calibrated only in the cycling-dominated regime (0–45 °C); and residual correlations of EOL damage with operating factors (0.280) and log lifetime (0.218) indicate incomplete stress-index coverage. Fourth, prediction intervals require the post-hoc variance inflation of Section IV-B (factor 2.887) to attain near-nominal coverage [44].

V. CONCLUSION

This study presents a physics-informed GPR framework for the probabilistic estimation of lithium-ion battery SOH. The framework integrates interspersed cyclic aging with EIS measurements. Impedance spectra are processed through a sequential LinKK and DRT validation pipeline to extract ECM parameters, which are combined with the CSI encoding operating history. On 256 unseen simulated cells the framework achieves 0.169% MAPE ($R^2 = 0.995$); on 94 cells outside the training distribution the error is 2.66% ($R^2 = 0.752$), attributable entirely to the cold tier, with the remaining extrapolation tiers pooling to the holdout figure. A single scalar variance recalibration attains 94.3 to 96.1% coverage on the primary partitions and reduces holdout NLPD from 1.517 to −0.825.

Kernel structure is consequential in this configuration: an additive impedance/stress covariance preserves a positive marginal contribution from the stress covariates, which a joint ARD kernel reverses, and the dependence of per-cell error on distance to the nearest training anchor, significant under the joint kernel, is not detectable under the additive one. CSI is largely redundant in distribution, yet predicts SOH standalone at $R^2 = 0.938$ on the holdout partition without any impedance measurement.

DATA AND CODE AVAILABILITY

The simulation configuration, the frozen model and calibration artifacts, the equivalent-circuit and Gaussian-process fitting code, and the complete parameter record (including all fitting thresholds, kernel bounds, and detection settings, each tagged with the basis on which it is set) are available from the corresponding author on reasonable request. The measured dataset is publicly available from [13].

REFERENCES

- [1] G. Saldaña, J. I. San Martín, I. Zamora, F. J. Asensio, O. Oñederra, and M. González-Pérez, “Empirical calendar ageing model for electric vehicles and energy storage systems batteries,” *J. Energy Storage*, vol. 55, 2022, Art. no. 105676.
- [2] M. Dabbarry, M. Berecibar, A. Devie, D. Anseán, N. Omar, and I. Villarreal, “State of health battery estimator enabling degradation diagnosis: Model and algorithm description,” *J. Power Sources*, vol. 360, pp. 59–69, 2017.
- [3] A. Irshayid and J. Chen, “Battery state-of-health estimation based on partial discharge curves,” in *Proc. IEEE Conf. Control Technol. Appl.*, 2026, pp. 67–72.
- [4] L. Nuculaj and J. Chen, “Dense extended kalman filter for simultaneous cell state-parameter estimation,” in *Proc. Amer. Control Conf.*, 2026, pp. 193–198.
- [5] B. Boukamp, “Electrochemical impedance spectroscopy in solid state ionics: Recent advances,” *Solid State Ionics*, vol. 169, no. 1–4, pp. 65–73, 2004.
- [6] W. Choi, H.-C. Shin, J. M. Kim, J.-Y. Choi, and W.-S. Yoon, “Modeling and applications of electrochemical impedance spectroscopy (eis) for lithium-ion batteries,” *J. Electrochem. Sci. Technol.*, vol. 11, no. 1, pp. 1–13, 2020.
- [7] Z. Szekeres, G. Márton, and G. Láng, “Methods for the determination of valid impedance spectra in non-stationary systems,” *ChemElectroChem*, vol. 8, no. 1, pp. 11–27, 2021.
- [8] J. Peng, Y. Gao, L. Cai, M. Zhang, C. Sun, and H. Liu, “State of health estimation for lithium-ion batteries using electrochemical impedance spectroscopy and a multi-scale kernel extreme learning machine,” *World Electric Veh. J.*, vol. 16, no. 4, 2025, Art. no. 224.
- [9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
- [10] R. R. Richardson, M. A. Osborne, and D. A. Howey, “Gaussian process regression for forecasting battery state of health,” *J. Power Sources*, vol. 357, pp. 209–219, Jul. 2017.
- [11] H. Chen, G. Dong, Y. Wang, J. Yu, L. Wu, and Y. Lou, “Data-driven battery health prognosis using scalable deep recurrent structure and partial fast-charging profiles,” *IEEE Trans. Veh. Technol.*, vol. 74, no. 11, pp. 17034–17046, Nov. 2025.
- [12] F. Li, Y. Min, Y. Zhang, Y. Zhang, H. Zuo, and F. Bai, “State-of-health estimation method for fast-charging lithium-ion batteries based on stacking ensemble sparse gaussian process regression,” *Rel. Eng. System Saf.*, vol. 242, 2024, Art. no. 109787.
- [13] Y. Zhang, Q. Tang, Y. Zhang, J. Wang, U. Stimming, and Y.-C. Lee, “Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning,” *Nature Commun.*, vol. 11, no. 1, 2020, Art. no. 1706.
- [14] M. Doyle, T. F. Fuller, and J. Newman, “Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell,” *J. Electrochem. Soc.*, vol. 140, no. 6, pp. 1526–1533, 1993.
- [15] V. Sulzer, S. G. Marquis, R. Timms, M. Robinson, and S. J. Chapman, “Python battery mathematical modelling (pybamm),” *J. Open Res. Softw.*, vol. 9, no. 1, 2021, Art. no. 14.
- [16] C.-H. Chen, F. Brosa Planella, K. O’Regan, D. Gastol, W. D. Widanage, and E. Kendrick, “Development of experimental techniques for parameterization of multi-scale lithium-ion battery models,” *J. Electrochem. Soc.*, vol. 167, no. 8, 2020, Art. no. 080534.
- [17] S. E. J. O’Kane et al., “Lithium-ion battery degradation: How to model it,” *Phys. Chem. Chem. Phys.*, vol. 24, no. 13, pp. 7909–7922, 2022.
- [18] R. Deshpande, M. Verbrugge, Y.-T. Cheng, J. Wang, and P. Liu, “Battery cycle life prediction with coupled chemical degradation and fatigue mechanics,” *J. Electrochem. Soc.*, vol. 159, pp. A1730–A1738, 2012.
- [19] M. A. Miner, “Cumulative damage in fatigue,” *J. Appl. Mechan.*, vol. 12, no. 3, pp. A159–A164, 1945.
- [20] P. C. Paris and F. Erdogan, “A critical analysis of crack propagation laws,” *J. Basic Eng.*, vol. 85, no. 4, pp. 528–533, 1963.
- [21] A. J. Torregrosa, A. Broatch, P. Olmeda, and L. Agizza, “A semi-empirical model of the calendar ageing of lithium-ion batteries aimed at automotive and deep-space applications,” *J. Energy Storage*, vol. 80, 2024, Art. no. 110388.
- [22] I. Baghdadi, O. Briat, J.-Y. Deléage, P. Gyan, and J.-M. Vlassas, “Lithium battery aging model based on dakin’s degradation approach,” *J. Power Sources*, vol. 325, pp. 273–285, Sep. 2016.

- [23] E. Redondo-Iglesias, P. Venet, and S. Pelissier, "Calendar and cycling ageing combination of batteries in electric vehicles," *Microelectronics Rel.*, vol. 88–90, pp. 1212–1215, 2018.
- [24] A. Karger et al., "Modeling particle versus SEI cracking in lithium-ion battery degradation: Why calendar and cycle aging cannot simply be added," *J. Electrochem. Soc.*, vol. 171, 2024, Art. no. 090512.
- [25] V. R. Joseph, E. Gul, and S. Ba, "Maximum projection designs for computer experiments," *Biometrika*, vol. 102, no. 2, pp. 371–380, 2015.
- [26] R. Li, S. O'Kane, J. Huang, M. Marinescu, and G. J. Offer, "A million cycles in a day: Enabling high-throughput computing of lithium-ion battery degradation with physics-based models," *J. Power Sources*, vol. 598, Apr. 2024, Art. no. 234184.
- [27] T. Waldmann, M. Wilka, M. Kasper, M. Fleischhammer, and M. Wohlfahrt-Mehrens, "Temperature dependent ageing mechanisms in lithium-ion batteries—a post-mortem study," *J. Power Sources*, vol. 262, pp. 129–135, 2014.
- [28] G. Kucinskis, M. Bozorgchenani, M. Feinauer, M. Kasper, M. Wohlfahrt-Mehrens, and T. Waldmann, "Arrhenius plots for Li-ion battery ageing as a function of temperature, C-rate, and ageing state—an experimental study," *J. Power Sources*, vol. 549, 2022, Art. no. 232129.
- [29] J. Vetter et al., "Ageing mechanisms in lithium-ion batteries," *J. Power Sources*, vol. 147, no. 1, pp. 269–281, 2005.
- [30] P. Keil et al., "Calendar aging of lithium-ion batteries," *J. Electrochem. Soc.*, vol. 163, no. 9, 2016, Art. no. A1872.
- [31] M. Ecker et al., "Calendar and cycle life study of li(nimnco)o₂-based 18650 lithium-ion batteries," *J. Power Sources*, vol. 248, pp. 839–851, 2014.
- [32] Z. Gao, H. Xie, X. Yang, W. Niu, S. Li, and S. Chen, "The dilemma of c-rate and cycle life for lithium-ion batteries under low temperature fast charging," *Batteries*, vol. 8, no. 11, Nov. 2022, Art. no. 234.
- [33] V. A. Agubra, J. W. Fergus, R. Fu, and S.-Y. Choe, "Analysis of effects of the state of charge on the formation and growth of the deposit layer on graphite electrode of pouch type lithium ion polymer batteries," *J. Power Sources*, vol. 270, pp. 213–220, Dec. 2014.
- [34] Y. Zheng et al., "Effects of state of charge on the degradation of lifepo₄/graphite batteries during accelerated storage test," *J. Alloys Compounds*, vol. 639, pp. 406–414, Aug. 2015.
- [35] M. Xu, X. Wang, L. Zhang, and P. Zhao, "Comparison of the effect of linear and two-step fast charging protocols on degradation of lithium ion batteries," *Energy*, vol. 227, Jul. 2021, Art. no. 120417.
- [36] M. Zhuo, G. Offer, and M. Marinescu, "Degradation model of high-nickel positive electrodes: Effects of loss of active material and cyclable lithium on capacity fade," *J. Power Sources*, vol. 556, Feb. 2023, Art. no. 232461.
- [37] B. Boukamp, "A linear kronig-kramers transform test for immittance data validation," *J. Electrochem. Soc.*, vol. 142, pp. 1885–1894, 1995.
- [38] F. Ciucci and C. Chen, "Analysis of electrochemical impedance spectroscopy data using the distribution of relaxation times: A bayesian and hierarchical bayesian approach," *Electrochimica Acta*, vol. 167, pp. 439–454, 2015.
- [39] B. Boukamp, "Distribution (function) of relaxation times, successor to complex nonlinear least squares analysis of electrochemical impedance spectroscopy," *J. Phys. : Energy*, vol. 2, 2020, Art. no. 042001.
- [40] B. Py and F. Ciucci, "The distribution of capacitive times for the analysis of electrochemical impedance spectroscopy data from super-capacitors," in *Proc. ECS Meeting Abstr.*, 2024, pp. 743–743.
- [41] B. Boukamp, "A nonlinear least squares fit procedure for analysis of immittance data of electrochemical systems," *Solid State Ionics*, vol. 20, pp. 31–44, 1986.
- [42] O. Orringer, "Rapid estimation of spectrum crack-growth life based on the palmgren-miner rule," *Comput. Structures*, vol. 19, pp. 149–153, 1984.
- [43] A. Marrel and B. Iooss, "Probabilistic surrogate modeling by gaussian process: A review on recent insights in learning and validation," *Rel. Eng. System Saf.*, vol. 247, 2024, Art. no. 110094.
- [44] B. Larvaron, M. Clausel, A. Bertonecello, S. Benjamin, and G. Oppenheim, "Chained gaussian processes to estimate battery health degradation with uncertainties," *J. Energy Storage*, vol. 67, Sep. 2023, Art. no. 107443.