Multi-level Constraints in PINNs: Theory and Applications to Battery Impedance Modeling

Ryuto Tanigawa, Yingrui Geng, Hayata Kaneko, Ryuto Ishibashi, Qi Li, Lin Meng

Ritsumeikan University, 1-1-1 Nojihigashi, Kusatsu, Shiga, Japan 525-8577

{ri0141ev, gr0731rp, ri0099xx, ri0097fx}@ed.ritsumei.ac.jp, {liqi24, menglin}@fc.ritsumei.ac.jp

Abstract

Accurate estimation of battery state-of-health (SOH) from electrochemical impedance spectroscopy (EIS) is crucial for safe and long-lived battery operation. Purely data-driven models, however, often struggle under limited, noisy, or out-of-distribution data. Hybrid physics–data models, including physics-informed neural networks (PINNs), can enhance robustness by incorporating physical priors; yet, systematic guidance on which priors are most beneficial, when, and why remains limited. We introduce a practical taxonomy that classifies priors into system-level constraints (S), which encode parametric, mechanism-level structure (e.g., equivalent-circuit models, Butler–Volmer relations, PDEs), and general-consistency constraints (G), which enforce model-agnostic validity (e.g., Kramers–Kronig relations, monotonicity). Using a constrained-hypothesis-space framework, we analyze how S and G affect identifiability and generalization, and we derive, under mild assumptions and compatibility conditions, an expected ordering of performance across constraint combinations. We also identify characteristic failure modes arising from overconstrained or incompatible priors. Empirically, we validate our theoretical insights using a public EIS dataset, evaluating SOH prediction with RMSE and MAE. Our experiments also evaluate several constraint combinations that have not been explored in prior PINN-based battery studies, enabling a systematic comparison across multi-prior configurations. Ablation and sensitivity studies further illustrate how prior informativeness, misspecification, and optimization error influence outcomes. Finally, we provide practical guidelines for selecting and combining priors, along with detailed experimental protocols to support reproducibility.

Keywords: Physics-Informed Neural Network, Lithium-ion battery, State of health estimation, Electrochemical impedance spectroscopy

1 Introduction

Accurate estimation of battery State of Health (SOH) is critical for the safe and efficient operation of battery systems. Electrochemical Impedance Spectroscopy (EIS) provides frequency-domain signatures associated with charge transfer, double-layer effects, and mass transport, and is widely used for diagnostics. Purely data-driven models can predict SOH from EIS, yet often lack interpretability and generalize poorly with limited or noisy data [1].

IIKI 2025, December 18-21, Chengdu, People's Republic of China

DOI 10.4108/eai.18-12-2025.2365280

Hybrid physics–data models incorporate physical priors to improve robustness and interpretability. However, prior work notes that the selection and integration of physical knowledge in such models are often handcrafted and task-specific, lacking systematic guidelines for what kind of physics to include and how to embed it effectively [2]. This manual and case-dependent design process complicates consistent comparison across studies and hinders scalable deployment. To address this ambiguity and to enable a more principled comparison of existing approaches, we propose a conceptual framework that categorizes physical priors into two practical classes: system-level (S) constraints that embed mechanistic, parameterized descriptions (e.g., equivalent-circuit models (ECM), Butler–Volmer kinetics (BV), partial differential equations (PDE)), and general-consistency (G) constraints that enforce model-agnostic physical validity (e.g., Kramers–Kronig relations (KK), monotonicity).

We analyze these constraints using a constrained-hypothesis-space formalism. Under mild realizability and regularity assumptions (that the constrained hypothesis class contains a plausible approximation of the true process and that the loss and constraint mappings are sufficiently smooth), we derive an ordering of expected performance across constraint combinations. Intuitively, S-type priors reduce hypothesis volume and improve identifiability and sample efficiency, whereas G-type priors act as softer regularizers that enhance physical plausibility while remaining more sensitive to approximation and optimization errors.

To validate these insights, we conduct experiments on a public EIS dataset measured at 25, 35, and 45°C. We compare hybrid architectures that combine system-level priors (ECM, BV, PDE), both individually and jointly, with general-consistency constraints including KK and monotonicity. Performance is evaluated primarily by RMSE and MAE under two data-size settings and three temperature conditions. Ablation studies isolate each constraint's contribution to overall performance. These empirical results support the theoretically predicted hierarchy and highlight trade-offs between prior informativeness and misspecification risk.

The key contributions of this paper are as follows:

- A conceptual constrained-hypothesis-space framework that clearly distinguishes between system-level (S) and general-consistency (G) constraints.
- A theoretical analysis under interpretable assumptions, highlighting the advantages of compatible system-level priors and the typical failure modes of mixed priors.
- A systematic empirical validation on a public EIS dataset, comparing multiple hybrid configurations using RMSE and MAE metrics.
- Practical guidelines for selecting and combining priors in EIS-based SOH estimation tasks.

2 Related Work

Hybrid physics–data frameworks are central to battery SOH estimation. In what follows, we review prior work through the lens introduced in the Introduction, distinguishing between mechanis-

tic, system-level constraints and model-agnostic, general-consistency constraints, and we highlight where existing studies fall short of a systematic comparison.

A large body of work embeds governing electrochemical laws directly into learning objectives or network architectures [3]. PINNs and related surrogates are applied to continuum electrochemical models to accelerate parameter inference and enable diagnostics; notable examples include PINN-based surrogates for single-particle (SPM) and pseudo-two-dimensional (P2D/DFN) models that speed up calibration and state estimation [4, 5]. Related efforts incorporate kinetic relations such as Butler–Volmer laws [6] or explicit ageing models [7, 8] into training to impose mechanistic behavior and improve interpretability. Another strand of research focuses on architectural priors: designing network modules that mirror equivalent-circuit computations or other mechanistic components yields models that are both more interpretable and more data-efficient than black-box counterparts [9, 10]. These studies collectively represent the system-level constraints defined in this work.

EIS naturally motivates a complementary class of constraints that are model-agnostic. Kramers–Kronig relations, which link the real and imaginary parts of impedance to enforce causality and passivity, are a canonical example; enforcing KK consistency is used as a diagnostic and a regularization tool in some EIS studies [11, 12, 13]. Likewise, simple physical plausibility conditions such as monotonicity of degradation indicators serve as lightweight, general-purpose regularizers [14, 15]. Practical challenges arise, however, when these integral or inequality constraints are applied to noisy, discrete data: numerical quadrature, sampling irregularities, and finite-sample noise can all undermine stable enforcement of such general-consistency priors.

Despite this rich literature, most prior studies focus on a single class of constraint and evaluate it in isolation. What remains largely missing is a principled, comparative perspective that characterizes how different types of priors interact, when they are complementary, and when combinations can harm optimization or increase approximation error. In particular, numerical stability and optimization difficulty—especially for heterogeneous mixes of integral and parametric constraints—are often treated as implementation details rather than as objects of formal analysis. This paper addresses these gaps by formalizing system-level and general-consistency priors within a constrained-hypothesis-space view, analyzing their bias–variance tradeoffs, and empirically evaluating their interactions on a common EIS benchmark.

3 Theoretical Framework

3.1 Overview

We analyze the effect of introducing multiple physics-based constraints into PINNs from a bias–variance perspective. Let $\mathscr{X}$ and $\mathscr{Y}$ denote the input and output spaces, respectively. The parametric hypothesis family is

$$\mathscr{H} = \{f_\theta : \theta \in \Theta\}, \qquad f_\theta : \mathscr{X} \to \mathscr{Y}, \tag{1}$$

where Θ is the parameter space and θ denotes model parameters. For a pointwise loss $\ell : \mathscr{Y} \times \mathscr{Y} \to \mathbb{R}_{\geq 0}$ we define the population (expected) risk

$$R(f) = \mathbb{E}_{(x,y)\sim\mathscr{D}}[\ell(f(x), y)] \tag{2}$$

and the empirical risk on an i.i.d. sample $(x_i, y_i)_{i=1}^n$,

$$\widehat{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i), \tag{3}$$

where n denotes the number of training examples and $\mathscr{D}$ the data-generating distribution. These quantities measure true expected loss and its finite-sample estimate, and serve as the basis for our generalization statements below.

3.2 Physics-constrained hypothesis classes

We focus on two practically useful categories of physics-based constraints. System-level (S) priors encode mechanistic structure (for example, equivalent-circuit models, Butler–Volmer kinetics, or PDE surrogates). To formalize these, let $C_S : \mathscr{H} \times \Phi \to \mathscr{V}_S$ be a constraint operator and Φ an auxiliary parameter space (e.g., ECM parameters). The feasible class satisfying the S constraint is

$$\mathscr{H}_S = \{f \in \mathscr{H} \mid \exists \phi \in \Phi :\ C_S(f;\phi) = 0\}. \tag{4}$$

General-consistency (G) priors impose model-agnostic validity conditions such as Kramers–Kronig relations, passivity, or monotonicity. Writing $C_G : \mathscr{H} \to \mathscr{V}_G$ for the corresponding operator, the feasible class for G is

$$\mathscr{H}_G = \{f \in \mathscr{H} \mid C_G(f) = 0\}. \tag{5}$$

When multiple priors are active, the effective hypothesis class is the intersection of the corresponding feasible sets; for example $\mathscr{H}_{S+S'} = \mathscr{H}_S \cap \mathscr{H}_{S'}$ or $\mathscr{H}_{S+G} = \mathscr{H}_S \cap \mathscr{H}_G$.

3.3 Bias–variance trade-off

Imposing constraints reduces the flexibility of the learning family and thus can lower estimation variance, while such constraints may also increase approximation bias if the priors exclude the true mapping. To quantify this trade-off we use a standard excess-risk bound [16]. Let $\mathscr{A} \subseteq \mathscr{H}$ denote any (possibly constrained) hypothesis class and $\hat{f}_{\mathscr{A}}$ its empirical risk minimizer (ERM) on n samples. Further let $\mathfrak{R}_n(\mathscr{A})$ denote an effective capacity measure of $\mathscr{A}$ (e.g., a Rademacher-type complexity), $B(\mathscr{A})$ the approximation bias of $\mathscr{A}$ relative to the true data-generating map f^*, and $\delta \in (0,1)$ a confidence parameter. Then with high probability

$$R(\hat{f}_{\mathscr{A}}) - R(f^*) \le 2\mathfrak{R}_n(\mathscr{A}) + B(\mathscr{A}) + c\sqrt{\frac{\log(1/\delta)}{n}}, \tag{6}$$

where c is a positive constant. The practical takeaway is immediate: adding a compatible, informative constraint should reduce the capacity term $\mathfrak{R}_n(\cdot)$ without materially increasing $B(\cdot)$, thereby tightening the bound; conversely, a misspecified constraint that substantially raises $B(\cdot)$ can negate or reverse any benefit.

3.4 Hierarchical constraint ordering (practical view)

When the true mapping satisfies multiple compatible system-level priors (i.e., $f^* \in \mathscr{H}_{S+S'}$) and the combined feasible class does not increase capacity, the intersection class $\mathscr{H}_{S+S'}$ inherits no larger excess-risk upper bound than either single-prior class. Equivalently, mutually consistent S-type priors can jointly reduce uncertainty and improve identifiability. By contrast, heterogeneous or numerically unstable combinations (for example, integral constraints with poor discretization) may introduce optimization difficulties and increase bias; such cases therefore call for careful discretization, conservative enforcement, or staged scheduling in training.

4 Methodology

4.1 Constraint overview

We implement two types of priors introduced in Sec. 3: system-level (S) priors, including ECM reconstruction, BV consistency, and PDE-based mass-transport residuals; and general-consistency (G) priors, including KK and monotonicity. The following subsections provide concrete discrete forms, the loss terms used during training, and optimization details for both soft-penalty and augmented-Lagrangian enforcement.

4.2 Training objectives

We implement constraints either as soft penalties or, when near-hard enforcement is desired, via an augmented-Lagrangian. Here θ denotes model parameters and ϕ auxiliary physical parameters (e.g., ECM variables).

$$\mathscr{L}_{\text{pen}}(\theta,\phi) = \mathscr{L}_{\text{data}}(\theta) + \lambda_S \|C_S(f_\theta;\phi)\|_{w_S}^2 + \lambda_G \|C_G(f_\theta)\|_{w_G}^2, \tag{7}$$

where $\|\cdot\|_w$ denotes initial-scale normalization (based on the magnitude of the initial impedance) and λ_S, λ_G are scalar penalty weights.

4.3 Individual constraint losses

Our hybrid model f_θ comprises multiple network branches (Fig. 1). A primary branch predicts the target SOH $\hat{y}$ from the input EIS spectrum x. Secondary branches predict intermediate physical parameters ϕ (e.g., R_{ct}, i_0, D_b) or the full spectrum Z^{pred}, which are used to compute the physical constraint losses. Below are the concrete loss components used in experiments (batch size B, linearized frequency points N_f).

$$L_{\text{data}} = \frac{1}{B}\sum_{b=1}^{B} (\hat{y}_b - y_b)^2, \tag{8}$$

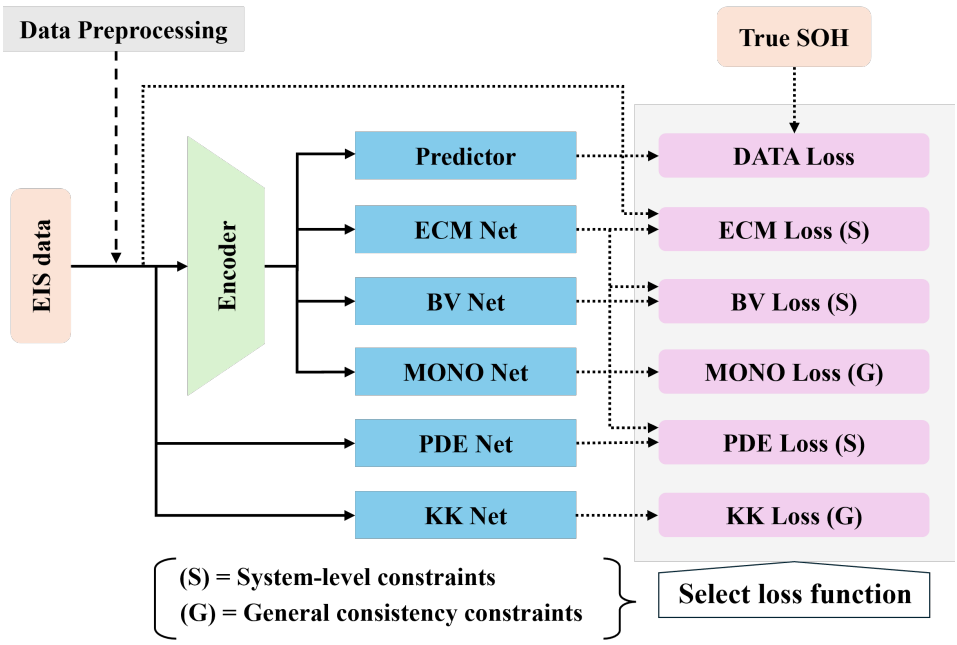


Fig. 1. Experimental hybrid architecture used to validate theoretical claims. All modules (Encoder, Predictor, ECM/BV/MONO/PDE/KK/Net) are implemented as MLPs. The model maps EIS inputs to SOH while imposing physics priors: system-level (S: ECM, Butler–Volmer, PDE) via explicit residuals or an augmented-Lagrangian, and general-consistency (G: Kramers–Kronig, monotonicity) via differentiable penalties or residuals depending on enforcement strength.

mean squared error on SOH predictions, where $\hat{y}_b = f_\theta(x_b)$ is the model's primary output for sample b.

$$L_{\text{ecm}} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_f}\sum_{f=1}^{N_f}\Big[(\Re Z_{b,f}^{\text{pred}} - \Re Z_{b,f}^{\text{obs}})^2 + (\Im Z_{b,f}^{\text{pred}} - \Im Z_{b,f}^{\text{obs}})^2\Big], \tag{9}$$

ECM reconstruction error (real and imaginary parts).

$$L_{\text{kk}} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_f}\sum_{f=1}^{N_f}\big(\widehat{\Re Z}[\Im Z^{\text{pred}}]_{b,f} - \Re Z_{b,f}^{\text{obs}}\big)^2, \tag{10}$$

KK (Hilbert or learned-map) mismatch for the real part.

Define the PDE residual $r_b(x,t) = c_t - D_b\, c_{xx}$ (here x,t denote spatial and temporal collocation points). Then

$$L_{\text{pde}} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{E}_{x,t}\left[r_b(x,t)^2\right] + |w_{\text{bc}}|\,\tfrac{1}{2}\left(\mathbb{E}_t[c_x(0,t)^2] + \mathbb{E}_t[c_x(1,t)^2]\right), \tag{11}$$

interior residual plus boundary-condition penalty (collocation averages); w_{bc} denotes the boundary-condition penalty weight.

$$R_{ct}^{\text{from}\, i_0} = \frac{RT}{nF\,(i_0+\varepsilon)}, \tag{12}$$

$$L_{\text{bv}} = \frac{1}{B}\sum_{b=1}^{B}\Big(\log(R_{ct,b}+\varepsilon) - \log(R_{ct,b}^{\text{from}\, i_0}+\varepsilon)\Big)^2, \tag{13}$$

Butler–Volmer consistency is implemented as the log-MSE between R_{ct} and the value implied by i_0, where R denotes the gas constant, T the temperature, n the number of transferred electrons, and F the Faraday constant. A small positive constant ε is added for numerical stability.

$$L_{\text{mono}} = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{N_m-1}\sum_{j=1}^{N_m-1}\big(\max(0,\, \hat{m}_{b,j+1} - \hat{m}_{b,j})\big)^2, \tag{14}$$

penalty for violations of monotonic SOH decrease over cycles. Here N_m is the number of sequential measurements/cycles and $\hat{m}_{b,j}$ denotes the MONO Net's predicted SOH for sample b at step j.

Finally, the weighted total loss used in experiments is

$$L_{\text{total}} = \sum_k \lambda_k L_k, \quad k \in \{\text{data}, \text{ecm}, \text{kk}, \text{pde}, \text{bv}, \text{mono}\}, \tag{15}$$

where a selected subset of $\{L_k\}$ is activated depending on the constraint configuration used in each experiment; each λ_k is a scalar weight.

4.4 Augmented Lagrangian formulation

For near-hard enforcement we use an augmented Lagrangian with initial-scale-normalized losses:

$$\mathscr{L}_{\text{AL}}(\theta,\phi,\mu) = \tilde{\mathscr{L}}_{\text{data}}(\theta) + \mu^\top \tilde{C}(\theta,\phi) + \frac{\rho}{2}\|\tilde{C}(\theta,\phi)\|_2^2, \tag{16}$$

where all terms are normalized by their initial scales (recorded at the beginning of training). Here, $\tilde{\mathscr{L}}_{\text{data}}$ and $\tilde{C}(\theta,\phi)$ denote the data loss and the concatenated residuals from active physics constraints (e.g., KK, PDE, BV, monotonicity, ECM), each divided by its initial magnitude. After inner updates of (θ,ϕ), multipliers are updated as $\mu \leftarrow \mu + \rho\,\tilde{C}(\theta,\phi)$ (using `detach()` for stability). We initialize $\mu = 0$ and start with a small ρ (the AL penalty parameter), optionally annealing it upward when residuals stagnate.

4.5 Implementation notes

In practice, physics-based constraints require careful scaling and scheduling. All residuals are normalized by their initial magnitudes to ensure comparable weights across experiments. Penalty terms are gradually increased, or updated via an augmented-Lagrangian scheme for those that are difficult to optimize, in order to avoid instability. Monitoring constraint residuals, multipliers, and gradient norms provides early indications of incompatibility or over-regularization. Overall, compatible system-level priors (S) yield the most stable and data-efficient learning, whereas general-consistency priors (G) act as flexible regularizers that are best introduced progressively.

5 Experiments

Table 1: Comparison of SOH estimation across loss combinations at different temperatures (mean ± std). Results are averaged over 30 random seeds per temperature. Lower values indicate better performance.

Combination of losses	25°C		35°C		45°C		Category
	MAE	RMSE	MAE	RMSE	MAE	RMSE	
DATA (Baseline)	0.0528±0.0052	0.0652±0.0053	0.0880±0.0426	0.1340±0.0571	0.1893±0.1111	0.2276±0.1306	Baseline
DATA + ECM	0.0531±0.0045	0.0659±0.0045	0.0752±0.0395	0.1031±0.0641	0.1166±0.0583	0.1409±0.0680	S_1
DATA + BV	0.0512±0.0049	0.0634±0.0049	0.0842±0.0428	0.1278±0.0610	0.1443±0.0908	0.1762±0.1051	S_2
DATA + PDE	0.0521±0.0042	0.0643±0.0042	0.0840±0.0310	0.1296±0.0498	0.1793±0.1194	0.2196±0.1423	S_3
DATA + ECM + BV	**0.0506**±0.0051	**0.0626**±0.0037	0.0656±0.0423	0.0960±0.0619	0.1054±0.0512	0.1292±0.0592	S_1+S_2
DATA + ECM + PDE	0.0520±0.0049	0.0639±0.0050	**0.0531**±0.0404	**0.0752**±0.0627	0.1166±0.0617	0.1413±0.0728	S_1+S_3
DATA + BV + PDE	0.0531±0.0069	0.0651±0.0064	0.0772±0.0283	0.1182±0.0467	0.1412±0.1038	0.1716±0.1244	S_2+S_3
DATA + KK	0.0522±0.0058	0.0645±0.0052	0.0886±0.0313	0.1320±0.0510	0.1745±0.0763	0.2126±0.0925	G_1
DATA + MONO	0.0532±0.0053	0.0651±0.0049	0.0788±0.0249	0.1226±0.0374	0.1809±0.0995	0.2177±0.1188	G_2
DATA + ECM + KK	0.0533±0.0094	0.0645±0.0091	0.0638±0.0541	0.0892±0.0740	0.1134±0.0604	0.1385±0.0739	S_1+G_1
DATA + ECM + MONO	0.0579±0.0156	0.0703±0.0201	0.1119±0.0471	0.1816±0.0757	0.1084±0.0633	0.1461±0.0703	S_1+G_2
(+Augmented Lagrangian)	0.0621±0.0108	0.0789±0.0097	0.0623±0.0106	0.0768±0.0134	**0.0631**±0.0089	**0.0741**±0.0100	S_1+G_2
DATA + KK + MONO	0.0512±0.0043	0.0636±0.0041	0.0947±0.0400	0.1426±0.0676	0.1771±0.0976	0.2146±0.1174	G_1+G_2

Notes: Values represent mean ± standard deviation calculated over $n = 30$ independent runs for each temperature condition. "System-level" refers to physics-based constraints (S_1: ECM, S_2: Butler–Volmer, S_3: PDE residuals), and "General-consistency" refers to domain-agnostic priors (G_1: Kramers–Kronig, G_2: Monotonicity). The best performance (lowest MAE and RMSE) for each temperature is highlighted in **boldface**.

5.1 Experimental Setup

5.1.1 Implementation Details

All experiments are conducted in Python 3.10.18 using PyTorch 2.7.1 (CUDA 12.6). The hardware platform consisted of an NVIDIA GeForce RTX 3080 Ti GPU (12 GB VRAM) running Ubuntu 22.04 with driver version 535.247.01.

5.1.2 Dataset

We use the publicly available electrochemical impedance spectroscopy (EIS) dataset accompanying Zhang *et al.* [17]. The dataset contains approximately 20,000 impedance spectra, each recorded at 60 discrete frequencies with real and imaginary parts stored separately (yielding 120 input features per spectrum). The sampled frequency range spans 0.02 Hz to 20 kHz. Measurements were taken on commercial lithium-ion cells ($LiCoO_2$/graphite) across multiple states of health and states of charge, and at three nominal temperatures: 25°C, 35°C and 45°C. Raw EIS spectra and accompanying capacity files are available on Zenodo (DOI: 10.5281/zenodo.3633835) and are distributed under a CC-BY-4.0 license.

5.1.3 Training Conditions

We use state-V data (measured after full charge with a 15-min rest). Each sample comprises 120 EIS features (60 real + 60 imaginary) and the corresponding cycle SOH. Cells 25C01–25C04, 35C01, and 45C01 form the training set, while 25C05–25C08, 35C02, and 45C02 are held out for testing. Twenty percent of the training samples is used for validation. All spectra are normalized by min–max scaling.

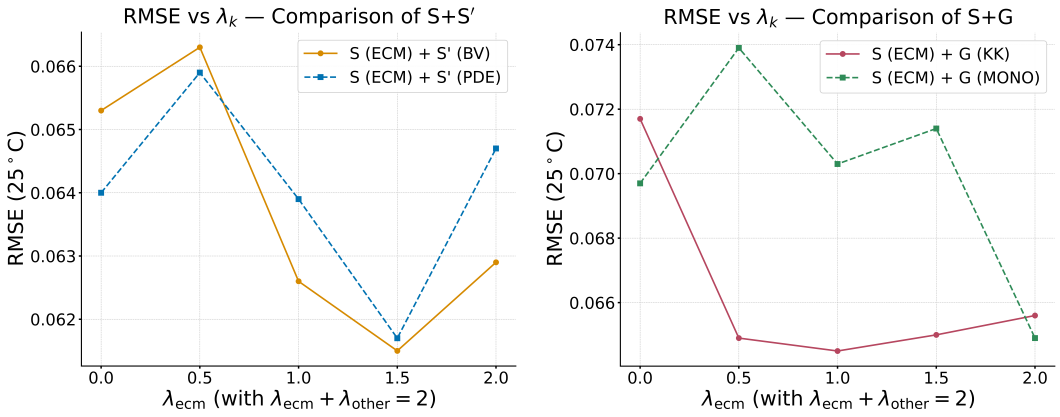
Models are trained with the Adam optimizer (batch size 64). The learning rate schedule uses a 30-epoch warmup (warmup lr = 1.5×10^{-3}) that linearly increases to the base learning rate (1×10^{-2}), followed by cosine decay from 1×10^{-2} to 2×10^{-4} over the remaining epochs (i.e., $200 - 30 = 170$ epochs); the cosine schedule is implemented so that the final rate equals 2×10^{-4} exactly. Early stopping is applied with patience = 30. The total loss is given in Eq. (15); in each run, the data loss L_{data} is combined with one or two physics-based constraint terms (selected from $\{\text{ecm}, \text{kk}, \text{pde}, \text{bv}, \text{mono}\}$). All weights are fixed at $\lambda_k = 1$, and each loss term is normalized at initialization to ensure balanced contributions. Each configuration is repeated 30 times with different random seeds for each temperature (25°C, 35°C, 45°C), yielding 90 independent runs.

5.1.4 Evaluation Metrics

We assess model performance using two regression metrics—Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE)—together with their corresponding standard deviations (MAE_std, RMSE_std).

5.2 Results

Table 1 summarizes MAE and RMSE (mean $\pm$ std) for SOH estimation across constraint combinations at 25°C, 35°C, and 45°C (30 seeds per entry). Mechanistic system-level priors (ECM, Butler–Volmer, PDE surrogates) consistently improve predictive accuracy over the data-only baseline, and compatible pairings of such priors (S+S′) yield the lowest mean errors and variance. Specifically, DATA+ECM+BV achieves the smallest errors at 25°C (MAE = 0.0506 $\pm$ 0.0051, RMSE = 0.0626 $\pm$ 0.0037) and performs strongly at 45°C, while DATA+ECM+PDE gives the best mean performance at 35°C.



(a) S+S′. **(b)** S+G.

Fig. 2. Ablation study: RMSE of SOH prediction (25°C) versus the ECM penalty λ_{ecm} (with $\lambda_{\text{ecm}} + \lambda_{\text{other}} = 2$). Left panel (a): system-level pairings (ECM+BV, ECM+PDE; denoted S+S′). Right panel (b): system–general pairings (ECM+KK, ECM+MONO; denoted S+G).

General-consistency priors (KK, monotonicity) offer modest gains and generally underperform compared to system-level priors. Introducing the augmented-Lagrangian optimization improves the difficult-to-optimize DATA+ECM+MONO at 35°C and 45°C, where data are sparse and noisy, while slightly deteriorating performance at 25°C, where abundant and stable data already enable accurate estimation. This indicates that the augmented-Lagrangian acts as an effective regularizer under data-sparse conditions, although it can introduce mild bias in well-conditioned regimes.

Fig. 2 further illustrates an ablation at 25°C, showing how RMSE varies with the ECM penalty weight. System-level pairings (S+S′) exhibit clearer minima and greater robustness than mixed S+G setups, which show higher variance due to weaker compatibility.

6 Discussion

6.1 Theoretical interpretation

The empirical results align with the constrained-hypothesis-space formalism in Sec. 3. Eq. (6) separates excess risk into (i) an estimation term governed by the capacity measure $2\mathfrak{R}_n(\mathscr{A})$, (ii) an approximation bias $B(\mathscr{A})$, and (iii) a confidence term $O(\sqrt{\log(1/\delta)/n})$. A prior that reduces $\mathfrak{R}_n$ without materially increasing $B(\cdot)$ improves generalization. This explains why compatible, mechanism-informed constraints reduce both error and variance, whereas incompatible or poorly scaled ones tend to increase bias or optimization error.

6.2 System-level priors

Mechanistic S priors (ECM, Butler–Volmer, PDE residuals) encode domain knowledge that directly constrains the mapping from spectra to SOH. When the true mapping f^* lies within the constrained class, intersections such as $\mathscr{H}_{S+S'}$ reduce degrees of freedom while preserving representational capacity, yielding the empirical relation S+S$'$ $>$ S, S$'$ (i.e., $R^*(\mathscr{H}_{S+S'}) \leq R^*(\mathscr{H}_S), R^*(\mathscr{H}_{S'})$ under compatibility). These priors guide representation learning toward physically meaningful latent variables (e.g., R_{ct}) and spectral reconstruction, leading to a reduction in effective complexity when such features correlate with SOH. Their roles are distinct yet cooperative: L_{ecm} promotes structural reconstruction, L_{bv} enforces kinetic consistency, and L_{pde} links the surrogate to transport physics—accounting for the observation that ECM+PDE achieves strong performance at 35°C, whereas BV provides dominant effects at higher temperatures.

6.3 General-consistency priors

General-consistency (G) priors, such as KK and monotonicity, act as plausibility regularizers that discourage physically inconsistent predictions without embedding mechanism-specific structure. Consequently, their impact on $\mathfrak{R}_n$ tends to be weaker and strongly dependent on numerical treatment. Discrete KK enforcement is sensitive to sampling and noise; the selection of quadrature weights w_{ij} and regularizer $\mathscr{R}_i$ determines whether KK acts to stabilize or amplify noise. Monotonicity priors improve plausibility, although they can become brittle if applied too aggressively or without proper normalization. Overall, G priors remain valuable components whose numerical behavior requires careful calibration.

6.4 Optimization, diagnostics, and statistical reliability

Mixed S+G configurations show two typical failure modes: (1) simultaneous enforcement with equal penalty scaling increases optimization difficulty, causing oscillatory multipliers and larger variance (e.g., DATA+ECM+MONO); (2) incompatible priors increase approximation bias $B(\cdot)$ and can invert expected performance orderings.

Three controls are most effective: (i) *initial-scale normalization* of constraint residuals to equalize loss scales; (ii) staged enforcement—ramping penalties from weak to strong—or the use of an augmented-Lagrangian (Eq. (16)) with monitored multiplier updates; and (iii) treating λ_k as learnable parameters (not used here, although considered promising).

A rapid growth of Lagrange multipliers indicates a mismatch in scaling or an incompatibility among constraints. Normalization is implemented by recording each residual at training start and dividing subsequent values by that magnitude, ensuring consistent interpretation of λ_k.

Statistically, large seed variance indicates multi-modal loss landscapes where initialization or SGD noise determines the final basin. Reporting both mean and standard deviation is therefore essential for reliability. Overall, mechanistic priors yield consistent gains when compatible and properly scaled, while consistency priors demand careful numerical handling. Normalization, staged enforcement, and multiplier monitoring effectively connect theoretical insight to empirical robustness.

7 Conclusion

This study investigates the impact of system-level (S) and general-consistency (G) priors on state-of-health (SOH) estimation from electrical impedance spectroscopy. Our findings indicate that compatible S priors effectively reduce hypothesis complexity and enhance identifiability, consistently yielding the highest predictive accuracy and minimizing variance. While G priors provide valuable plausibility regularization, their efficacy heavily depends on careful discretization and rigorous penalty management. Combining multiple compatible system-level priors (S+S′) typically tightens excess-risk bounds and improves empirical performance. Conversely, heterogeneous mixtures demand cautious, staged enforcement and active optimization diagnostics to prevent performance degradation. In practice, researchers should first verify S-prior compatibility, apply initial-scale normalization, tune penalty scaling, and introduce constraints progressively—such as transitioning from soft starts to stronger penalties or employing augmented-Lagrangian methods. Continuous monitoring of data loss, per-constraint residuals, and multiplier magnitudes is crucial to ensure robust and interpretable hybrid estimation. Ultimately, enhancing the reliability of SOH predictions through these physics-informed methodologies will contribute significantly to the development of safer, more efficient battery management systems for next-generation energy storage applications. We anticipate this framework will serve as a foundational guide for further research on the design, validation, and optimal selection of constraint combinations in physics-informed machine learning models for battery diagnostics and beyond.

References

[1] Akram AS, Sohaib M, Choi W. SOH Estimation of Lithium-ion Batteries using LSTM Model with Deconvoluted EIS Parameters. In: 2025 IEEE Energy Conversion Congress & Exposition Asia (ECCE-Asia). IEEE; 2025. p. 1-6.

[2] Meng C, Griesemer S, Cao D, Seo S, Liu Y. When physics meets machine learning: A survey of physics-informed machine learning. Machine Learning for Computational Science and Engineering. 2025;1(1):20.

[3] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics. 2019;378:686-707.

[4] Hassanaly M, Weddle PJ, King RN, De S, Doostan A, Randall CR, et al. PINN surrogate of Li-ion battery models for parameter inference, Part II: Regularization and application of the pseudo-2D model. Journal of Energy Storage. 2024;98:113104.

[5] Wang F, Zhai Z, Zhao Z, Di Y, Chen X. Physics-informed neural network for lithium-ion battery degradation stable modeling and prognosis. Nature Communications. 2024;15(1):4332.

[6] Dickinson EJ, Wain AJ. The Butler-Volmer equation in electrochemical theory: Origins, value, and practical application. Journal of Electroanalytical Chemistry. 2020;872:114145.

[7] Ye J, Xie Q, Lin M, Wu J. A method for estimating the state of health of lithium-ion batteries based on physics-informed neural network. Energy. 2024;294:130828.

[8] Tu H, Moura S, Wang Y, Fang H. Integrating physics-based modeling with machine learning for lithium-ion batteries. Applied Energy. 2023;329:120289.

[9] Wang F, Zhi Q, Zhao Z, Zhai Z, Liu Y, Xi H, et al. Inherently interpretable physics-informed neural network for battery modeling and prognosis. IEEE Transactions on Neural Networks and Learning Systems. 2023.

[10] Tran MK, Mathew M, Janhunen S, Panchal S, Raahemifar K, Fraser R, et al. A comprehensive equivalent circuit model for lithium-ion batteries, incorporating the effects of state of health, state of charge, and temperature on model parameters. Journal of Energy Storage. 2021;43:103252.

[11] Urquidi-Macdonald M, Real S, Macdonald DD. Applications of Kramers—Kronig transforms in the analysis of electrochemical impedance data—III. Stability and linearity. Electrochimica Acta. 1990;35(10):1559-66.

[12] Liu J, Wan TH, Ciucci F. A Bayesian view on the Hilbert transform and the Kramers-Kronig transform of electrochemical impedance data: Probabilistic estimates and quality scores. Electrochimica Acta. 2020;357:136864.

[13] Lazanas AC, Prodromidis MI. Electrochemical impedance spectroscopy- a tutorial. ACS measurement science au. 2023;3(3):162-93.

[14] Sánchez L, Anseán D, Otero J, Couso I. Assessing the health of LiFePO4 traction batteries through monotonic echo state networks. Sensors. 2017;18(1):9.

[15] Tanigawa R, Geng Y, Kaneko H, Meng L. A hybrid MLP-Attention physics-informed neural network for robust SOH estimation of lithium-ion batteries. In: 2025 Int'l Conference on Advanced Mechatronic Systems (ICAMechS). IEEE; 2025. p. 96-101.

[16] Bartlett PL, Mendelson S. Rademacher and gaussian complexities: Risk bounds and structural results. Journal of machine learning research. 2002;3(Nov):463-82.

[17] Zhang Y, Tang Q, Zhang Y, Wang J, Stimming U, Lee AA. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. Nature communications. 2020;11(1):1706.