energies

MDPI

*Article*

# Cross-Battery Lithium-Ion Battery SOH Estimation Using Reference-Normalized Local Discharge-Time Features

**Yong Chen, Quanhui Hou *, Jigang Liu and Yaodong Wang**

School of Automotive Engineering, Yancheng Institute of Technology, Yancheng 224051, China
* Correspondence: hqhdyx66@ycit.edu.cn

**Abstract**

Cross-battery state-of-health (SOH) estimation requires health indicators that remain informative when target-cell data are excluded from model training. We examine five fixed-voltage local discharge-time windows and three feature representations: raw time, initial-reference offset, and relative change. Four NASA cells (636 discharge records) and four CALCE CS2 cells (4011 records) are evaluated using battery-level protocols. The primary source-only nested leave-one-battery-out (LOBO) NASA experiment achieves a battery-macro RMSE of 0.01613 (between-cell SD 0.00349), with 3.7–3.5 V and 3.8–3.6 V each selected in two outer folds. In a separate retrospective fixed-window analysis at 3.7–3.5 V, relative normalization reduces NASA macro RMSE from 0.02104 to 0.01267, but the improvement on CALCE is marginal (0.02733 to 0.02719); offset subtraction yields 0.02684 on CALCE. Direct NASA-to-CALCE and CALCE-to-NASA cross-dataset transfers favor offset subtraction over relative scaling. These transfers do not involve retraining on target-dataset SOH labels but require an initial discharge-time reference from each target battery. Additional analyses examine reference-cycle exclusion, later references, deterministic reference perturbations, conventional single-feature baselines, and source-only support vector regression (SVR) tuning. These results support reference-assisted use under specific conditions rather than a universal advantage of relative normalization. The limited number of independent cells and dependence on a reliable target-cell reference constrain generalization.

**Keywords:** lithium-ion battery; state of health; health indicator; cross-battery generalization; reference normalization; leave-one-battery-out validation; partial discharge

## 1. Introduction

State of health (SOH) is commonly estimated from measurable electrical or thermal signals because capacity, the quantity used here to define SOH, is not continuously available during normal battery operation. Data-driven studies therefore construct health indicators (HIs) from voltage, current, temperature, or time and learn a mapping from these indicators to capacity-derived SOH [1–4]. These methods range from compact, physically interpretable indicators to multi-indicator fusion and sequence models [2–7].

A common objective is to use only part of a charge or discharge curve. Zhu et al. related constant-current charging time to SOH using linear regression, showing that a single time-domain indicator can be useful when its relationship with degradation is sufficiently consistent [2]. Yu et al. included equal-voltage-drop discharge time among several HIs [3], whereas Zhang et al. used discharge-derived temperature features with a temporal convolutional network [4]. More recent studies have examined partial charging segments,

Check for updates

Academic Editor: Jacob J. Lamb

Received: 23 September 2026
Revised: 21 October 2026
Accepted: 21 October 2026
Published: 22 October 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

sensitivity to the available voltage interval, and optimized charging windows [1,5,6,8]. Studies published in Energies have also shown that SOH can be estimated from partial-discharge features [9], enhanced indirect health indicators derived from cycling signals [10], and indirect voltage/current-based indicators when direct capacity measurement is impractical [11]. Thus, the use of local curve segments and voltage-window HIs is well established.

A central challenge is whether an HI derived from one group of cells remains predictive for unseen cells. Chen et al. explicitly formulate cross-cell SOH estimation as a domain-shift problem [12], and Yang et al. search for voltage intervals that generalize across cells rather than optimizing an interval independently for each cell [13]. Barua et al. showed on the same four NASA cells that random cycle splits can yield markedly different performance estimates from leave-one-cell-out testing [14]. Le and Nguyen reached a similar conclusion when comparing charging-phase HIs under LOBO validation [15]. Petkovski et al. also used a battery-independent test subset [9], while Yu et al. evaluated transfer among NASA random-walk cells and incorporated particle-filter updates when capacity observations became available [11]. Battery-level generalization has therefore been investigated previously; the present study evaluates reference-assisted discharge-time representations and their transfer limitations.

This study examines the discrepancy between within-cell aging correlation and cross-battery prediction error. Several neighboring discharge-time windows can track aging almost monotonically, yet their errors on an unseen battery may differ. We therefore investigate whether removing a cell-specific feature baseline before regression can reduce this discrepancy, without assuming that 3.7–3.5 V is universally optimal.

The proposed feature transforms require a valid initial local discharge-time measurement from the target battery. The method is therefore reference-assisted and does not provide reference-free diagnosis for a used battery with unknown history. We distinguish within-dataset validation (with entire batteries held out) from direct NASA-to-CALCE and CALCE-to-NASA cross-dataset transfer. In the latter, target-cell reference features are available, but target SOH labels are not used for retraining.

This study makes three main contributions: (i) a controlled ablation comparing raw, offset, and relative feature representations to separate baseline removal from division by the reference; (ii) a battery-level validation framework that distinguishes retrospective fixed-window comparisons from source-only nested selection and includes analyses of reference exclusion, reference timing, reference perturbations, and model tuning; and (iii) a comparison of within-dataset replication with reference-assisted cross-dataset transfer to assess practical applicability.

*Positioning Relative to Recent Literature*

Table 1 compares the present study with closely related work. Previous studies support compact time/voltage HIs, partial-curve feature extraction, optimized voltage intervals, and cross-cell testing [2,3,5,6,9,11,13]. The present analysis evaluates feature performance based on both within-cell aging associations and prediction error when an entire battery is held out. It also assesses whether normalization can mitigate cell-specific differences in feature baselines.

**Table 1.** Positioning of the present study relative to closely related SOH-estimation literature.

| Study | Feature/Data Focus | Validation Emphasis | What It Establishes | Difference from This Work |
|---|---|---|---|---|
| [2] Zhu et al., 2022 | CC charging time; NASA + CALCE | Within-dataset estimation | Single interpretable HI + LR can be effective | Charging-time HI; not evidence of LOBO transfer |
| [3] Yu et al., 2023 | Multi-dimensional HIs incl. equal-voltage-drop discharge time | Chronological split on NASA B0005 | Discharge-time HI is established prior art | Single-cell 100/18/50 cycle split; not LOBO |
| [4] Zhang et al., 2022 | Discharge-derived temperature variation rate; TCN | Prediction-start/sliding-window evaluation | Correlation-based HI extraction + TCN | NASA temporal evaluation; not demonstrated LOBO |
| [6] Wang et al., 2026 | Optimized charging interval + ensemble learning | Charging-interval transfer across nominal capacities | Voltage-interval optimization is established prior art | Transfer setting differs from NASA ↔ CALCE direct transfer |
| [9] Petkovski et al., 2024 | Partial-discharge features from voltage intervals; SVR | 109 training cells; 15 independent test cells | Partial-discharge voltage-range features are viable SOH inputs | Battery-held-out testing on 124 LFP cells; not four-fold LOBO |
| [10] Liu et al., 2020 | Enhanced indirect HI; NASA data; SVR | Per-cell chronological train/validation/test | Feature enhancement and SVR can support data-driven SOH estimation | First 90/next 30/later cycles; not cross-cell LOBO |
| [11] Yu et al., 2017 | Indirect voltage/current HIs under randomized use | Cross-battery testing; PF label updating | Measurable indirect HIs can replace direct capacity as model inputs | RW3–RW5 battery-held-out tests; PF may incorporate later target SOH labels |
| [13] Yang et al., 2026 | Cross-cell voltage interval mining + attention | Cross-cell evaluation on three datasets | Cross-cell interval mining and transferable interval selection | Multi-dataset testing does not alone establish direct cross-chemistry transfer |
| [14] Barua et al., 2026 | NASA predictor/validation study | Random-cycle split versus LOCO | Validation protocol can inflate apparent reliability | Highlights validation inflation and cell-transfer reliability |
| [15] Le & Nguyen, 2026 | CC/CV charging HIs | NASA leave-one-battery-out | Charging HI performance changes under cross-battery validation | Charging CC/CV indicators, not discharge-time features |
| [16] Zhuo et al., 2026 | Multi-indicator dependency model; NASA + CALCE | NASA/CALCE temporal model evaluation | Adaptive multi-indicator fusion can achieve high accuracy | Time-series tuning and multiple datasets do not imply direct transfer |

## 2. Materials and Methods

Figure 1 summarizes the complete workflow from public aging datasets and local discharge-time extraction to reference normalization and battery-level validation.

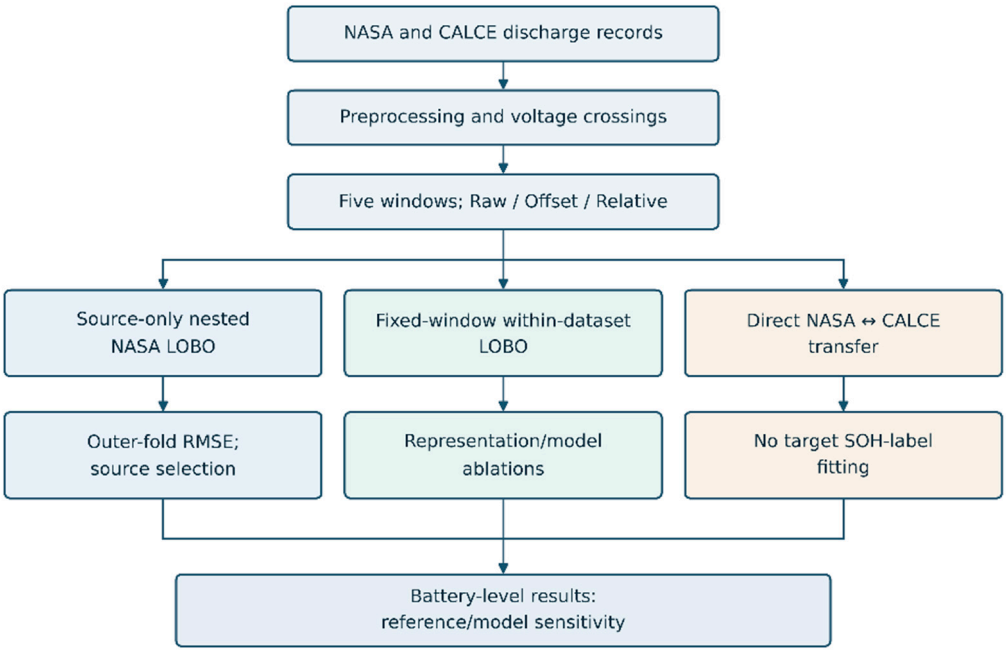



**Figure 1.** Workflow for reference-assisted SOH estimation using local discharge-time features. Three independent protocols are distinguished: source-only nested NASA LOBO, retrospective fixed-window within-dataset LOBO, and direct NASA–CALCE transfer without target SOH-label fitting. Target-battery initial references are required for transformed features.

### *2.1. Datasets and SOH Definition*

The main experiments use the NASA Prognostics Center of Excellence Battery Data Set [17], provided by the NASA Ames Research Center (Moffett Field, CA, USA). After parsing, cells B0005, B0006, B0007, and B0018 contribute 168, 168, 168, and 132 valid discharge records, respectively, for 636 records in total. For battery b at discharge cycle c, SOH is defined from the measured discharge capacity as follows:

$$SOH_b(c) = \frac{Q_b(c)}{Q_b(1)} \quad (1)$$

In Equation (1), $Q_b(1)$ is the first valid measured discharge capacity of battery b. SOH is therefore dimensionless and referenced to each cell's own initial measured capacity, rather than to rated capacity. The initial SOH value equals 1 by definition. Figure 2 shows the resulting SOH trajectories and the local discharge-time feature used in the main analysis. Capacity is used to construct the evaluation target; the feature transform in Equation (4) uses discharge time, not capacity. Recovering absolute capacity from a predicted SOH would additionally require $Q_b(1)$.

The replication dataset is the CALCE CS2 archive [18], provided by the Center for Advanced Life Cycle Engineering (CALCE), University of Maryland (College Park, MD, USA). CS2_35, CS2_36, CS2_37, and CS2_38 are LiCoO2 prismatic cells in the 1C-cycled group [18–21]. Excel files are ordered by the calendar date parsed from the filename. Within each file, cycles are processed in ascending Cycle_Index order. Samples with current below −0.5 A identify candidate discharge steps; the step containing the largest number of such samples is selected. At least five qualifying samples are required, and these are ordered by Test_Time(s). A discharge is retained only when the maximum voltage among its first five samples is at least 3.7 V, its minimum voltage is at most 2.8 V, and both selected threshold crossings yield a positive local time. Capacity is calculated as the maximum minus the minimum Discharge_Capacity(Ah) over the entire local cycle, rather than as the maximum

cumulative capacity alone; values must lie in (0, 2.0] Ah. Accepted cycles are assigned a continuous Global_Cycle index across files. This procedure yields 930, 970, 1036, and 1075 records, respectively (4011 total). The selected voltage window is not re-optimized on CALCE. Table S6 summarizes the retained samples and initial references.

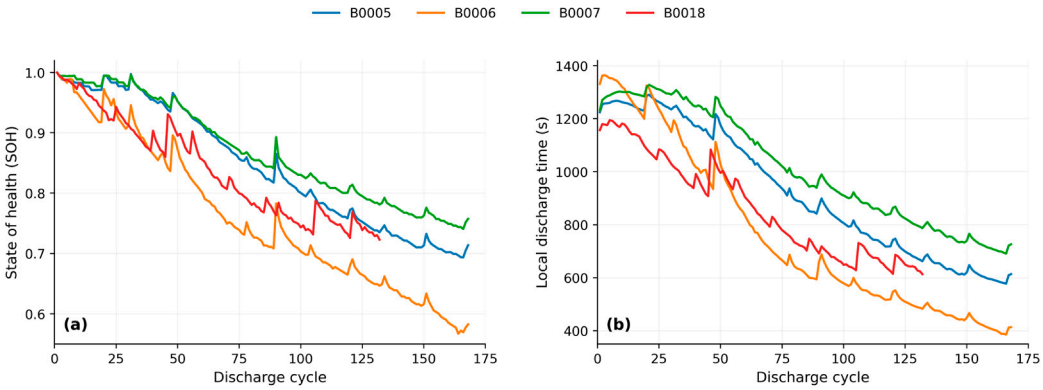



**Figure 2.** NASA battery degradation trajectories: (**a**) capacity-based state of health (SOH); (**b**) local discharge time obtained by linear interpolation between 3.7 and 3.5 V. SOH is defined relative to each cell's first valid measured discharge capacity, rather than its manufacturer-rated capacity.

### *2.2. Local Discharge-Time Health Indicator*

For each discharge trace, the first downward crossings of an upper threshold $V_u$ and a lower threshold $V_l$ are located. Because a threshold generally lies between two sampled voltage points ($t_j$,$V_j$) and ($t_{j+1}$,$V_{j+1}$), its crossing time at $V_*$ is obtained by linear interpolation using Equation (2):

$$t(V_*) = t_j + \frac{V_* - V_j}{V_{j+1} - V_j}(t_{j+1} - t_j) \quad (2)$$

The local discharge-time HI for cycle c is then defined by Equation (3) as follows:

$$T^c_{u-l} = t^c_l - t^c_u \quad (3)$$

In Equations (2) and (3), the subscripts u and l identify the upper and lower voltage thresholds; T is measured in seconds. The first pair of samples bracketing a downward threshold crossing is used. Equation (2) applies when the two voltages differ. For an equal-voltage plateau at the threshold, the implementation returns the earlier sample time; the NASA implementation also applies this fallback when the voltage difference is below $10^{-12}$ V. A local-time feature is retained only when both crossing times are finite, and the lower crossing occurs later than the upper crossing. The five candidate windows are 4.0–3.8, 3.9–3.7, 3.8–3.6, 3.7–3.5, and 3.6–3.4 V. For the multi-window comparison, a NASA record must be valid for all five windows. The resulting 636 records match the single-window dataset in battery, cycle, capacity, SOH, and 3.7–3.5 V feature values, ensuring that the same cycles are used to compare all candidates. Section 2.4 distinguishes the definition of candidate windows from their selection through validation.

### *2.3. Initial-Reference Normalization*

For each battery and voltage window, let $T_{(b,c)}$ be the interpolated local discharge time at cycle c and $Tref_{(b)}$ its first valid measured value. Three representations are compared: Raw = $T_{(b,c)}$; Offset = $T_{(b,c)} - Tref_{(b)}$; and Relative = $[T_{(b,c)} - Tref_{(b)}]/Tref_{(b)}$. A reference measurement from each target battery is required for Offset and Relative, but target SOH

labels are not used for training. Equation (4) gives the Relative representation; the Offset representation is included to isolate the effect of baseline subtraction.

$$R^{c}_{u-l} = \frac{T^{c}_{u-l} - T^{1}_{u-l}}{T^{1}_{u-l}} \quad (4)$$

The reference feature is computed independently for each battery and candidate voltage window. In the primary evaluation protocol, the reference record may be included among the records used for prediction. Additional sensitivity analyses exclude it from held-out evaluation only or from both training and evaluation while retaining the reference value solely for feature construction.

Reference-timing experiments also consider records 2, 3, and 5. These experiments distinguish evaluations restricted to observations after the chosen reference from other timing comparisons. Sensitivity is assessed using deterministic perturbations of ±1%, ±2%, and ±5% to the target reference; these perturbations do not represent a probabilistic sensor-noise model.

### 2.4. Battery-Level Validation and Nested Selection

NASA LOBO uses three complete source batteries for model fitting and one held-out battery for evaluation. The battery-macro metric is the arithmetic mean of four battery-level errors. The 3.7–3.5 V window was selected after inspecting the overall NASA LOBO results, so comparisons using this fixed window are retrospective. The primary nested analysis uses source-only voltage-window selection, as specified below. CALCE fixed-window LOBO provides within-dataset replication rather than direct transfer. Separate NASA-to-CALCE and CALCE-to-NASA experiments train on the source dataset without retraining on target SOH labels; target reference features remain available for Offset and Relative.

The controlled comparison of Raw, Offset, and Relative uses the same four outer NASA battery folds and an inner LOBO procedure involving the three source batteries. Within each outer fold, the voltage window minimizing the mean inner-battery RMSE is selected independently for each representation, using only the source batteries. The held-out battery is used exclusively for outer evaluation.

### 2.5. Regression Models and Metrics

Linear regression (LR) is the principal estimator. LR and support vector regression with a radial basis function kernel (RBF-SVR) use StandardScaler fitted on training data only. Random forest (RF) uses 300 trees, random_state = 42 and n_jobs = −1, without an explicit maximum-depth cap; gradient boosting regression (GBR) uses 200 estimators, learning_rate = 0.05, max_depth = 3, random_state = 42. The fixed-parameter RBF-SVR uses C = 10, epsilon = 0.01 and gamma = “scale”. A separate source-only nested SVR tuning analysis uses C = {0.1, 1, 10, 100}, epsilon = {0.001, 0.01, 0.05} and gamma = {“scale”, 0.1, 1.0}; selection excludes the held-out battery. These results do not constitute an exhaustive comparison of optimally tuned estimators.

For each held-out battery, MAE is the mean absolute prediction error, and RMSE is the square root of the mean squared prediction error. $R^2$ is one minus the ratio of the residual sum of squares to the total sum of squares about that battery’s mean observed SOH. Macro metrics are arithmetic means over the four batteries, and worst-cell RMSE is the maximum of the four battery-level RMSE values. Errors are expressed on the dimensionless SOH scale. Relative RMSE changes use the raw result as the denominator. Percentage changes are calculated from unrounded macro RMSE values and reported to two decimal places; displayed RMSE values are rounded separately. These percentage changes are descriptive effect sizes, not significance tests. Data preprocessing, feature extraction, and

model evaluation were performed using Python (version 3.12.14), NumPy (version 2.5.3), pandas (version 3.0.5), SciPy (version 1.18.1), and scikit-learn (version 1.9.1). Matplotlib (version 3.11.2) was used for data visualization.

## 3. Results

### *3.1. Aging Correlation Versus Cross-Battery Transferability*

The five predefined local discharge-time windows differ in their association with SOH, and their rankings by aging correlation do not necessarily match their rankings by cross-battery prediction error. The mean absolute Spearman correlation coefficients for the 4.0–3.8, 3.9–3.7, 3.8–3.6, 3.7–3.5, and 3.6–3.4 V windows are 0.97108, 0.99659, 0.99708, 0.99594, and 0.99556, respectively. The corresponding battery-macro LOBO RMSE values obtained using the raw features are 0.03082, 0.02061, 0.01981, 0.02104, and 0.02455. Among these candidates, the 4.0–3.8 V window has the lowest mean absolute correlation and the highest prediction error. The remaining four windows show similarly strong aging correlations but different cross-battery errors (Figure 3). These findings indicate that within-cell aging correlation alone is insufficient to determine whether a health indicator transfers to an unseen battery.

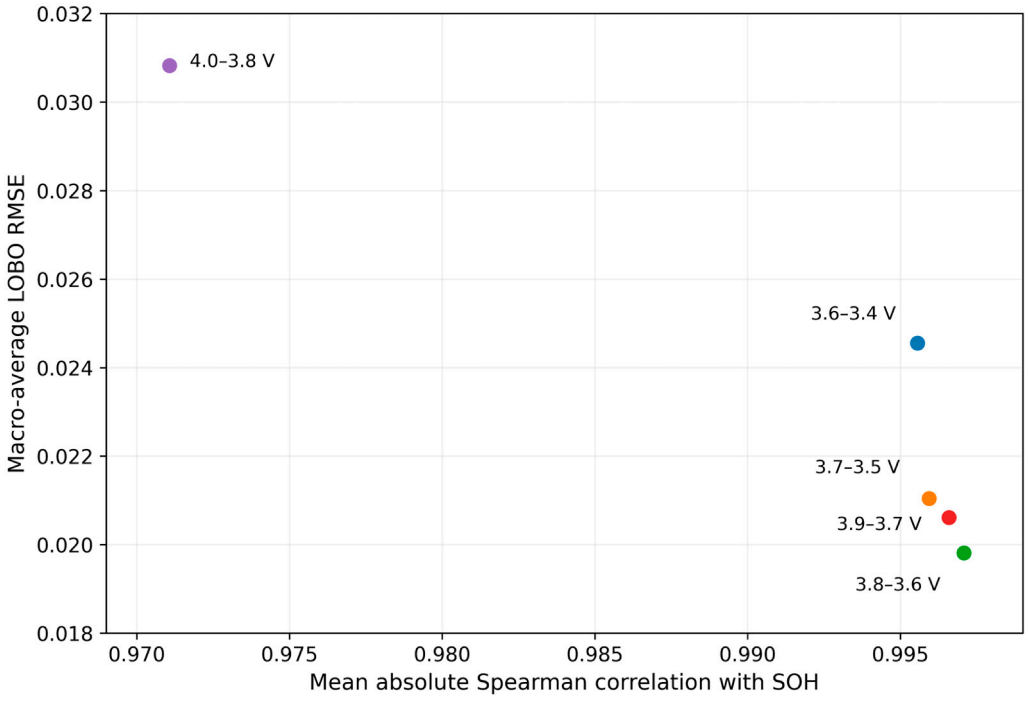



**Figure 3.** Aging correlation versus cross-battery transferability for the five predefined raw local discharge-time features. Similar within-cell aging correlation does not imply identical battery-held-out prediction error.

### *3.2. Reference Normalization on NASA*

For the selected fixed 3.7–3.5 V window, NASA LR macro MAE decreases from 0.01832 to 0.01063, and macro RMSE decreases from 0.02104 to 0.01267, a 39.76% reduction in macro RMSE. Macro $R^2$ increases from 0.94452 to 0.98156, while the largest cell-level RMSE decreases from 0.02934 to 0.02039. The cell-level RMSE pairs (raw, normalized) are B0005 (0.01110, 0.00883), B0006 (0.02794, 0.02039), B0007 (0.01576, 0.01570), and B0018 (0.02934, 0.00578). The improvements are uneven: B0007 improves only slightly, whereas B0018 shows the largest absolute reduction. Figure 4 shows the two feature representations, and

Figure 5 shows the normalized-feature predictions. These fixed-window results should be interpreted alongside the nested-selection results in Section 3.3.

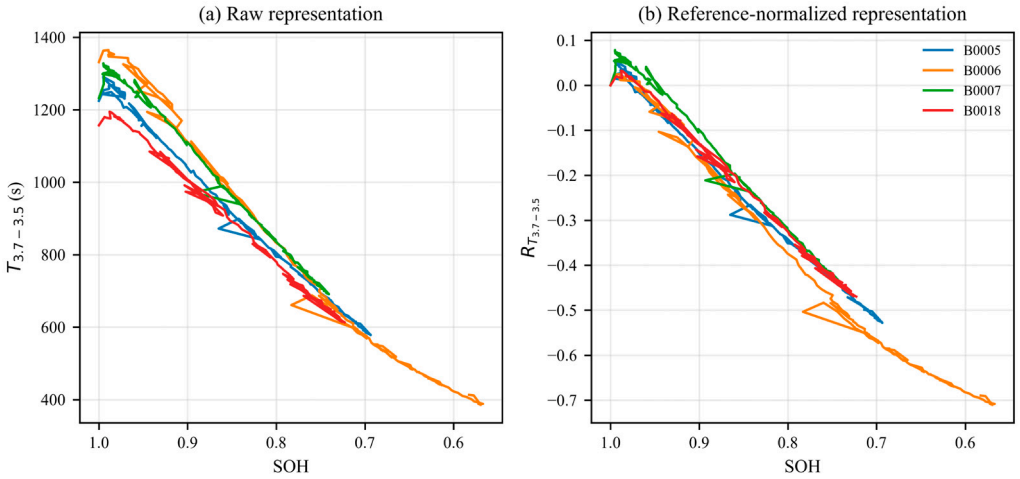



**Figure 4.** Raw and initial-reference-normalized local discharge-time representations in the 3.7–3.5 V window across four NASA cells: (**a**) raw local discharge time; (**b**) relative change normalized by each cell's initial reference value.

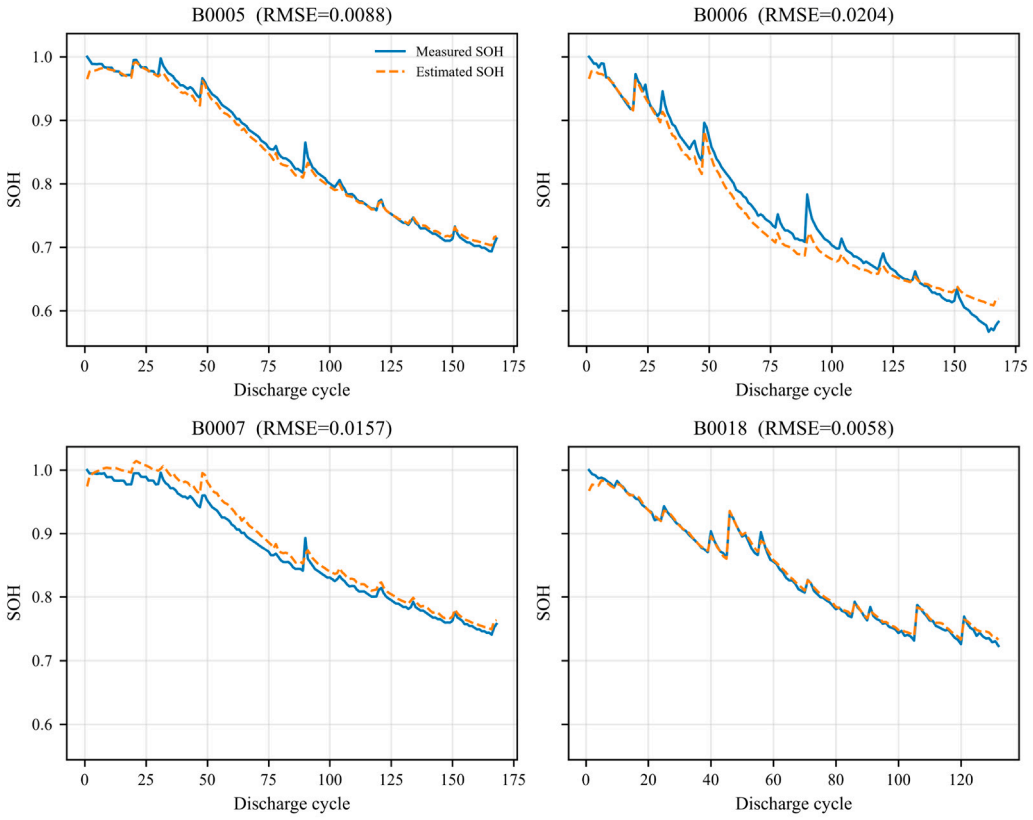



**Figure 5.** Measured and estimated SOH trajectories for the four NASA LOBO folds using linear regression and the normalized 3.7–3.5 V feature. RMSE values in subplot titles are rounded for display.

### *3.3. Strict Nested Selection*

For the Relative representation, source-only nested LOBO validation yields a macro outer-fold RMSE of 0.01613, a macro $R^2$ of 0.97297, and a maximum outer-fold RMSE of 0.02039. Using only source-battery data, the procedure selects the normalized 3.8–3.6 V

window when B0005 or B0018 is held out, and the normalized 3.7–3.5 V window when B0006 or B0007 is held out. This result is also included in the matched comparison of Raw, Offset, and Relative under the same nested protocol. It evaluates source-only voltage-window selection for Relative, whereas the RMSE of 0.01267 in Section 3.2 corresponds to the fixed candidate identified from the overall NASA comparison. The fixed-window and nested-selection results are reported separately because they reflect different validation procedures. The variation in selected windows suggests that a mid-voltage region is useful, but does not establish a universally optimal voltage window. Battery-level outer-fold results are presented in Figure 6.

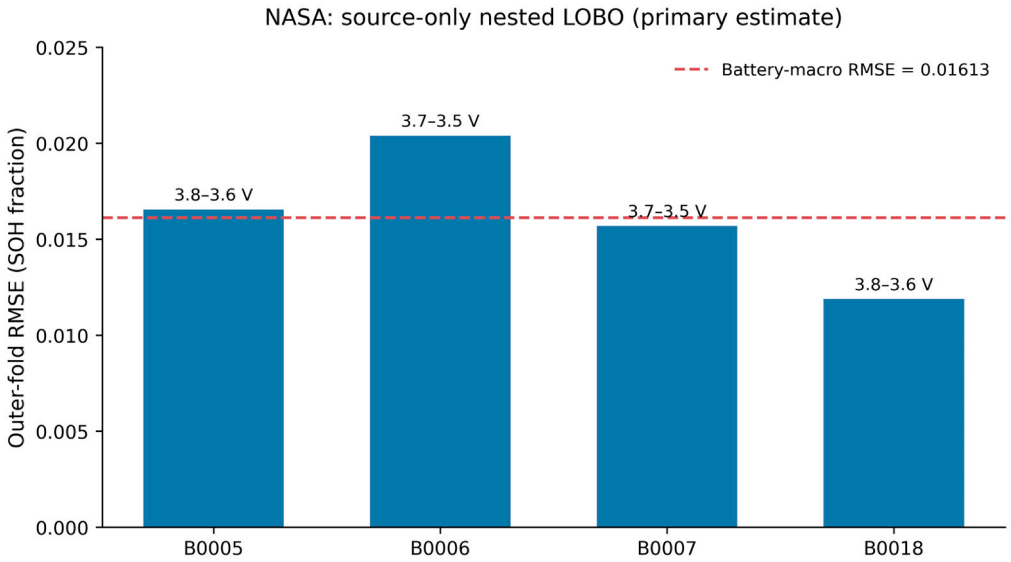



**Figure 6.** Held-out battery RMSE from source-only nested NASA LOBO. The feature configuration was selected using the three source batteries in each outer fold; the dashed line denotes battery-macro RMSE (0.01613).

The outer-fold RMSE values are 0.01655 for B0005, 0.02039 for B0006, 0.01570 for B0007, and 0.01189 for B0018. Table S2 reports the feature selected using only source batteries and the outer-fold RMSE for each held-out NASA battery. The inner loop leaves one of the three source batteries out for validation and trains on the other two; the outer held-out battery is excluded from this ranking.

Under identical source-only nested NASA LOBO procedures, representation-specific battery-macro RMSE values were 0.025148 (Raw), 0.020636 (Offset), and 0.016131 (Relative); worst-cell RMSE values were 0.035920, 0.036548, and 0.020388, respectively (Table 2). Relative had the lowest macro and worst-cell RMSE, but Offset had a slightly higher worst-cell RMSE than Raw. These comparisons do not imply superiority for every target battery or a universal advantage in cross-dataset transfer.

The representation-specific outer-fold windows and errors are provided in Supplementary Table S7.

### *3.4. Model Robustness*

The effect of fixed-window normalization depends on the regression model. RF macro RMSE falls from 0.02870 to 0.02126 (25.92%), and GBR macro RMSE falls from 0.02766 to 0.02024 (26.82%). The cell-level comparisons show lower RMSE for LR, RF, and GBR on all four NASA cells. RBF-SVR behaves differently: macro RMSE rises from 0.02070 to 0.02719 (31.34%), mainly because error increases in the B0006 fold. Figure 7 therefore supports a benefit of normalization for the tested linear and tree-based configurations but not across

all models. These comparisons keep the selected window fixed; nested window selection is not repeated separately for each regressor.

**Table 2.** Representation-specific source-only nested NASA LOBO with independent source-only window selection. RMSE is on the dimensionless SOH scale; macro RMSE is the unweighted mean of four target-cell RMSE values.

| Representation | Macro RMSE | Worst-Cell RMSE | Selected Windows by Held-Out Battery |
|---|---|---|---|
| Raw | 0.025148 | 0.035920 | B0005: 3.8–3.6 V, B0006: 3.9–3.7 V, B0007: 3.7–3.5 V, B0018: 3.7–3.5 V |
| Offset | 0.020636 | 0.036548 | B0005: 3.7–3.5 V, B0006: 3.7–3.5 V, B0007: 3.7–3.5 V, B0018: 3.7–3.5 V |
| Relative | 0.016131 | 0.020388 | B0005: 3.8–3.6 V, B0006: 3.7–3.5 V, B0007: 3.7–3.5 V, B0018: 3.8–3.6 V |

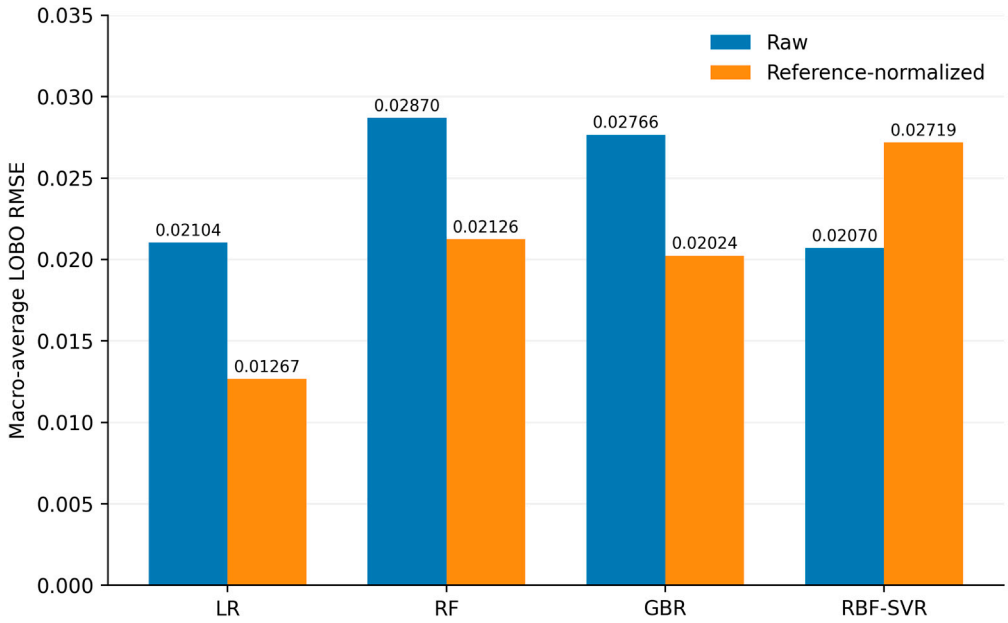



**Figure 7.** Model robustness comparison of raw and reference-normalized representations on NASA under battery-level LOBO. LR, RF, and GBR show lower macro RMSE after normalization, whereas the tested RBF-SVR configuration shows higher macro RMSE. These comparisons use fixed model configurations; source-only RBF-SVR tuning results are reported separately in Section 3.9.

The improvements across all cells are specific to RMSE. For LR on B0007, MAE rises from 0.01375 to 0.01410 despite the small RMSE decrease. Maximum absolute error on B0006 also rises from 0.06400 to 0.07068. Normalization therefore does not improve every error measure on every battery. Table S1 summarizes the macro-level model comparison. The unfavorable RBF-SVR result is specific to the fixed hyperparameter configuration; source-only hyperparameter tuning reverses the relative performance of Raw and Relative, as reported in Section 3.9.

### *3.5. Independent CALCE Replication*

The CALCE experiment keeps the 3.7–3.5 V feature definition fixed. Full-life within-CALCE LOBO gives macro RMSE values of 0.02733 for the raw feature and 0.02719 for the normalized feature, a 0.51% reduction. In the restricted analysis, the SOH $\geq$ 0.70 criterion is applied to both source and held-out batteries before a new model is fitted in

each fold. This retains 654, 641, 727, and 782 records for CS2_35, CS2_36, CS2_37, and CS2_38, respectively (2804 total), and gives macro RMSE values of 0.01993 and 0.01971 (1.10% reduction). Reference normalization still uses the first valid full-life record. Because SOH labels define the operating range, this analysis is retrospective and cannot be used to select unseen cycles when their SOH is unknown. Figure 8 compares the Raw, Offset, and Relative representations under fixed-window LOBO for NASA and CALCE. Table S1 provides the corresponding macro-level error summaries. No statistical significance is inferred from the small aggregate changes.

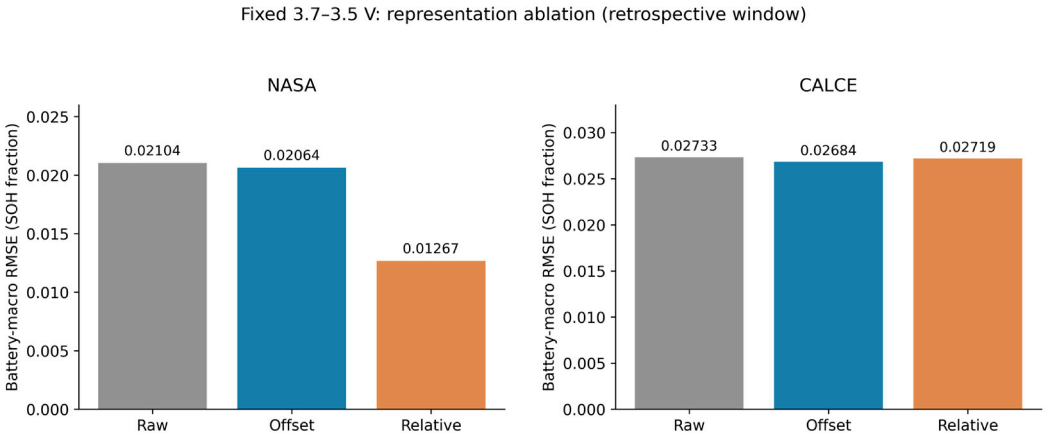



**Figure 8.** Raw, offset, and relative local discharge-time representations under fixed 3.7–3.5 V within-dataset LOBO linear regression. The fixed window was selected retrospectively using NASA results; these comparisons are descriptive and do not provide an unbiased estimate of window-selection performance.

### *3.6. Feature-Representation Ablation*

At the retrospectively selected 3.7–3.5 V window, NASA battery-macro RMSE values were 0.021037 for Raw, 0.020636 for Offset, and 0.012673 for Relative. The corresponding CALCE values were 0.027332, 0.026841, and 0.027191 (Figure 8). Relative achieved the largest improvement on NASA, whereas differences among the three representations were small on CALCE. Thus, subtracting the initial reference and scaling by that reference have different effects across the two datasets.

### *3.7. Direct Cross-Dataset Transfer*

Figure 9 reports direct cross-dataset transfer without fitting to target-dataset SOH labels. NASA-to-CALCE battery-macro RMSE was 0.238513 (Raw), 0.033711 (Offset), and 0.106896 (Relative); CALCE-to-NASA RMSE was 0.240325, 0.016308, and 0.085185. The transformed representations require an initial feature reference from each target battery. Offset performed best in both directions.

The target-cell results in Table 3 show that Offset attained the lowest RMSE for each of the four CALCE cells in NASA-to-CALCE transfer and each of the four NASA cells in CALCE-to-NASA transfer. However, errors vary by target cell: Offset RMSE is 0.025886–0.042374 for CALCE targets and 0.007959–0.037767 for NASA targets. These transfer results apply to the evaluated NASA–CALCE dataset pair and require an initial discharge-time reference from each target cell.

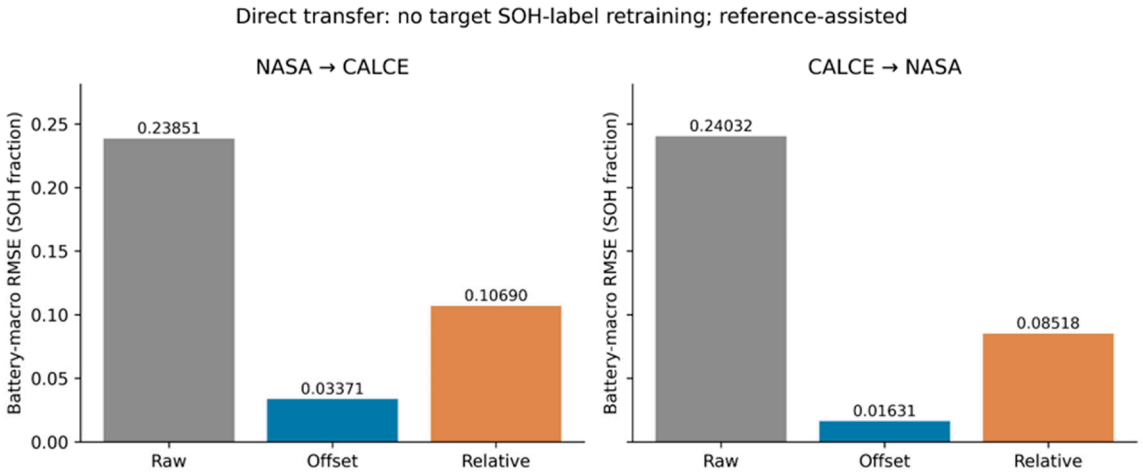



**Figure 9.** Direct NASA-to-CALCE and CALCE-to-NASA transfer without target SOH-label retraining. Offset and Relative use the target battery's initial local discharge-time reference; bars report battery-macro RMSE.

**Table 3.** Target-cell RMSE for direct cross-dataset transfer without target SOH-label retraining. Offset and Relative require each target cell's initial discharge-time reference.

| Direction | Target Cell | Raw | Offset | Relative |
|---|---|---|---|---|
| NASA → CALCE | CS2_35 | 0.236419 | 0.025886 | 0.095293 |
| NASA → CALCE | CS2_36 | 0.246095 | 0.042374 | 0.123806 |
| NASA → CALCE | CS2_37 | 0.238478 | 0.036291 | 0.109408 |
| NASA → CALCE | CS2_38 | 0.233061 | 0.030294 | 0.099077 |
| CALCE → NASA | B0005 | 0.244645 | 0.010740 | 0.080695 |
| CALCE → NASA | B0006 | 0.225260 | 0.037767 | 0.127058 |
| CALCE → NASA | B0007 | 0.229984 | 0.008765 | 0.059316 |
| CALCE → NASA | B0018 | 0.261410 | 0.007959 | 0.073670 |

Corresponding RMSE and MAE values for each target cell are reported in Supplementary Table S8.

### *3.8. Reference Exclusion, Timing, and Perturbation*

When the initial reference record is excluded from both model training and held-out evaluation, the LR Relative macro RMSE is 0.012335 for NASA and 0.027181 for CALCE (Table S1). Later reference records (2, 3, and 5) were evaluated under two distinct timing protocols, with results reported in Table S3. Deterministic target-reference perturbations of ±1%, ±2%, and ±5% were evaluated. At ±5%, the NASA Relative RMSE increases to 0.023542 (−5%) and 0.020778 (+5%), while the CALCE Relative RMSE increases to 0.040081 and 0.038060. Supplementary Figures S1 and S2 show the reference-perturbation and reference-timing analyses, respectively.

### *3.9. Model Sensitivity and Conventional Single-Feature Comparisons*

On NASA, the fixed-parameter RBF-SVR macro RMSE is 0.020704 (Raw) and 0.027193 (Relative); source-only tuning changes these values to 0.020066 and 0.014498. Six conventional single-feature LR baselines were evaluated using NASA battery-level LOBO. The mean voltage baseline gives an RMSE of 0.037588, and the total discharge time baseline gives an RMSE of 0.057691. Measurement requirements differ, and these simple baselines do not constitute a comprehensive reproduction of published methods. Supplementary Figure S3 presents the SVR tuning comparison.

### *3.10. Independent-Cell Uncertainty and Label Definition*

The nested NASA outer-fold RMSE values are 0.016553 for B0005, 0.020388 for B0006, 0.015696 for B0007, and 0.011885 for B0018; the four-cell mean is 0.016131, and the sample SD is 0.003489. Capacity-based SOH is defined relative to the first valid measured discharge capacity, not rated capacity. An exploratory nominal-capacity scenario is tabulated separately in the Supplementary Materials and does not replace the primary label definition.

### *3.11. Consolidated Results by Validation Protocol*

Table 4 summarizes the battery-macro RMSE results across the three evaluation protocols: retrospective fixed-window within-dataset LOBO, source-only nested NASA LOBO, and direct cross-dataset transfer.

**Table 4.** Battery-macro RMSE by validation protocol and feature representation.

| Protocol | Representation | Macro RMSE | Condition |
|---|---|---|---|
| NASA fixed-window LOBO | Raw | 0.021037 | Retrospective |
| NASA fixed-window LOBO | Offset | 0.020636 | Retrospective |
| NASA fixed-window LOBO | Relative | 0.012673 | Retrospective |
| CALCE fixed-window LOBO | Raw | 0.027332 | Fixed NASA window |
| CALCE fixed-window LOBO | Offset | 0.026841 | Fixed NASA window |
| CALCE fixed-window LOBO | Relative | 0.027191 | Fixed NASA window |
| NASA nested source-only LOBO | Relative (source-only selected window) | 0.016131 | Primary estimate |
| NASA→CALCE direct transfer | Raw | 0.238513 | No target-label fitting |
| NASA→CALCE direct transfer | Offset | 0.033711 | Target reference |
| NASA→CALCE direct transfer | Relative | 0.106896 | Target reference |
| CALCE→NASA direct transfer | Raw | 0.240325 | No target-label fitting |
| CALCE→NASA direct transfer | Offset | 0.016308 | Target reference |
| CALCE→NASA direct transfer | Relative | 0.085185 | Target reference |

Note: Nested source-only selection is the primary NASA generalization estimate. Fixed-window comparisons are retrospective. Direct transfer uses no target SOH labels for fitting; transformed features use target initial references.

## 4. Discussion

### *4.1. Correlation and Transferability Are Different HI Properties*

The window experiment distinguishes two properties that are often treated as interchangeable. Spearman correlation describes how consistently a feature changes with SOH within the observed aging trajectories. LOBO error assesses whether a mapping fitted on other batteries remains calibrated for the held-out battery. A cell-specific offset or scale can leave within-cell rank correlation almost unchanged while shifting the cross-cell regression relationship. This explains why the mid-voltage windows in Figure 3 can all have absolute Spearman correlations near 1 yet produce different held-out errors. The need to state the held-out unit explicitly is consistent with earlier battery-level test protocols [9,11] and more recent comparisons of cycle-level and cell-level validation [12,14,15].

### *4.2. Interpretation of the Normalization Effect*

The initial 3.7–3.5 V discharge-time features had coefficients of variation (CVs) of 5.818% for NASA and 0.518% for CALCE (Figure 10). Each CV was calculated from the four initial battery-level measurements as 100 times the sample standard deviation (ddof = 1) divided by their mean. The corresponding reference values are provided in Table S6.

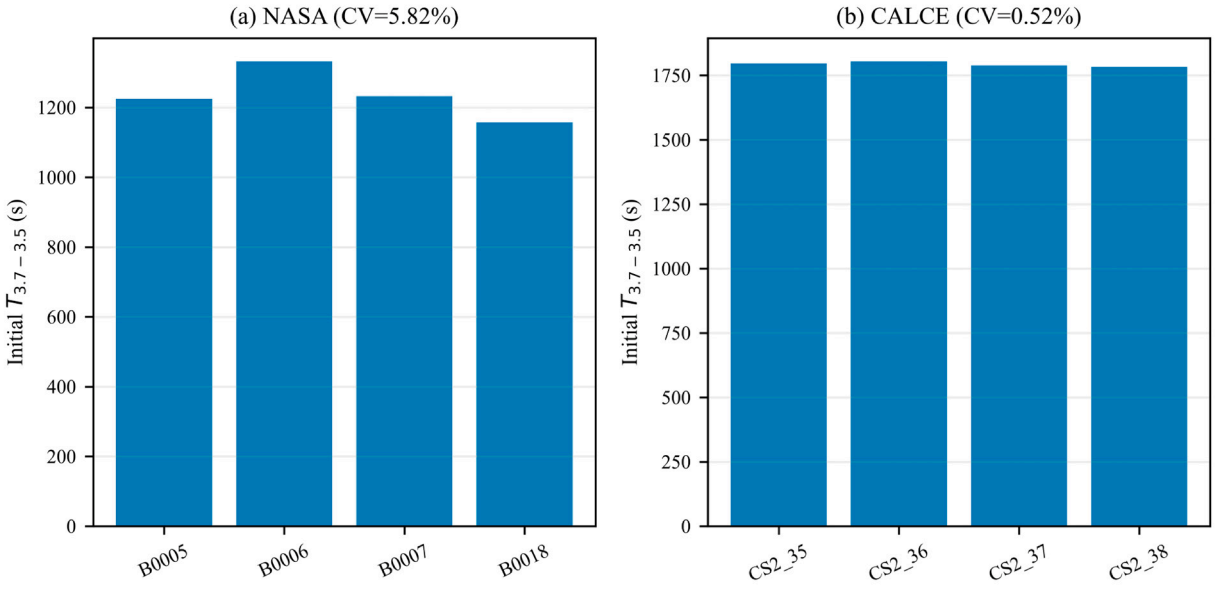



**Figure 10.** Initial 3.7–3.5 V local discharge-time feature baseline dispersion: (**a**) NASA dataset (CV = 5.82%); (**b**) CALCE dataset (CV = 0.52%). CV denotes the coefficient of variation. The corresponding unrounded values reported in the text are 5.818% and 0.518%, respectively.

The representation ablation showed that reference subtraction and scaling have different effects. Relative achieved the lowest macro RMSE in the source-only nested NASA evaluation (0.016131), whereas Offset yielded the lowest direct-transfer RMSE for all eight target cells. Offset removes the initial discharge-time baseline, while Relative also scales the difference by that baseline. Differences in initial-feature dispersion may contribute to the observed behavior, but NASA and CALCE also differ in cell characteristics, experimental protocols, and operating histories. The available comparisons do not establish baseline dispersion as the cause of the performance differences.

### *4.3. Why CALCE Shows Only Marginal Normalization Gain*

The CALCE results show that normalization does not consistently improve within-dataset accuracy. For the fixed 3.7–3.5 V window, macro RMSE is 0.027332 (Raw), 0.026841 (Offset), and 0.027191 (Relative). The Raw-to-Relative change is only approximately 0.51%. These results concern within-CALCE LOBO rather than direct NASA-to-CALCE transfer. The SOH $\geq$ 0.70 subset analysis is retrospective because it uses SOH labels to define the evaluation range.

### *4.4. Stable Mid-Voltage Region Rather than a Universal Optimal Threshold*

The predefined overlapping 0.2 V windows form an empirical, equal-width candidate set rather than thresholds uniquely determined by electrochemistry. Source-only nested selection chooses 3.7–3.5 V in two outer folds and 3.8–3.6 V in two. With only three source cells per outer fold, the evidence supports a potentially informative mid-voltage region, not a universal optimum. Under approximately constant current, local voltage-crossing time depends on charge throughput and voltage response, both of which are affected by capacity, resistance and polarization. The data cannot isolate a unique degradation mechanism.

### *4.5. Practical Scope and Limitations*

Offset and Relative require a reliable initial local discharge-time reference from each target battery across the selected voltage interval. Obtaining this measurement may require commissioning time, access to the relevant state-of-charge range, and comparable current

and temperature conditions. Feasibility for used cells without a suitable reference and application-specific acquisition costs remain unquantified.

The evaluation includes four NASA cells and four CALCE cells. The many cycles within each cell are correlated observations, not independent battery replicates. The primary nested NASA macro RMSE was 0.01613 with a between-cell SD of 0.00349. These descriptive statistics do not establish narrow population confidence intervals. The 39.76% improvement from the fixed-window NASA comparison is retrospective, whereas source-only nested selection provides the primary within-NASA generalization estimate.

CALCE fixed-window LOBO assesses within-dataset replication, whereas direct NASA-to-CALCE and CALCE-to-NASA tests assess reference-assisted transfer without retraining on target SOH labels. Neither protocol demonstrates reference-free deployment. Moreover, SOH relative to the first measured capacity differs from SOH defined using rated capacity. A valid local-time feature also requires both voltage thresholds to be crossed; variation in current, temperature, polarization, and sampling resolution may affect the measured feature. Additional chemistries and operating conditions remain to be tested.

Model behavior also depends on configuration. The prediction accuracy of fixed-parameter RBF-SVR decreased after Relative normalization, whereas source-only tuning reduced its Relative RMSE from 0.027193 to 0.014498, compared with 0.020066 for tuned Raw. This difference may reflect changes in kernel distances or source–target alignment, but the mechanism was not isolated. Future work should examine larger cell populations, varied duty cycles, the cost of acquiring reference measurements, reference-free alternatives, and robustness to measurement conditions. The ±1–5% deterministic reference perturbations reported here do not constitute a stochastic sensor-noise model.

## 5. Conclusions

The source-only nested NASA LOBO evaluation yielded a battery-macro RMSE of 0.01613 (SD 0.00349) for the Relative representation, compared with 0.025148 for Raw and 0.020636 for Offset under the same selection protocol. In the retrospective fixed-window analysis, Relative reduced NASA RMSE from 0.02104 to 0.01267, whereas its improvement on CALCE was marginal (0.02733 to 0.02719). Offset performed better than Relative in both direct cross-dataset transfer directions and achieved the lowest RMSE for all eight target batteries. These results indicate that the preferred discharge-time representation depends on the evaluation and transfer setting. The approach requires an initial target-battery discharge-time reference, and its applicability beyond the eight evaluated cells and their operating conditions remains to be established.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/en19215042/s1, Tables S1–S8 and Figures S1–S3. The supplementary materials provide additional experimental results, model parameters, reproducibility checks, and sensitivity analyses.

**Author Contributions:** Methodology, Y.C. and Q.H.; software, Y.C.; investigation, Q.H.; writing—original draft preparation, Y.C.; writing—review and editing, J.L., Y.W., and Q.H.; funding acquisition, Y.C. and Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the funding for school-level research projects of Yancheng Institute of Technology (Grant number: xjr2026026) and the Yancheng Institute of Technology Industry–University Cooperation Research Project (Grant number: YG20260824007).

**Data Availability Statement:** The NASA and CALCE battery aging datasets used in this study are publicly available from the NASA Prognostics Data Repository and the CALCE Battery Research Data archive, respectively, as cited in the References [17,18]. The processed data supporting the reported results are available from the corresponding author upon reasonable request, subject to applicable

dataset terms and institutional restrictions. The study-specific implementation code is not publicly available. The preprocessing procedures, feature definitions, model configurations, and validation protocols are described in the manuscript and Supplementary Materials to support methodological transparency.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References


1. Shu, X.; Yang, H.; Liu, X.; Feng, R.; Shen, J.; Hu, Y.; Chen, Z.; Tang, A. State of health estimation for lithium-ion batteries based on voltage segment and transformer. *J. Energy Storage* **2025**, *108*, 115200. [CrossRef]
2. Zhu, X.; Wang, W.; Zou, G.; Zhou, C.; Zou, H. State of health estimation of lithium-ion battery by removing model redundancy through aging mechanism. *J. Energy Storage* **2022**, *52*, 105018. [CrossRef]
3. Yu, Z.; Liu, N.; Zhang, Y.; Qi, L.; Li, R. Battery SOH Prediction Based on Multi-Dimensional Health Indicators. *Batteries* **2023**, *9*, 80. [CrossRef]
4. Zhang, D.; Zhao, W.; Wang, L.; Chang, X.; Li, X.; Wu, P. Evaluation of the State of Health of Lithium-Ion Battery Based on the Temporal Convolution Network. *Front. Energy Res.* **2022**, *10*, 929235. [CrossRef]
5. Vilsen, S.B.; Stroe, D.-I. Sensitivity analysis for reliable state-of-health estimation based on battery partial charging. *Cell Rep. Phys. Sci.* **2025**, *6*, 102646. [CrossRef]
6. Wang, B.; Cao, W.; Li, M.; Wang, J.; Zhao, F.; Yan, Y.; Wang, C. State of Health Estimation for Lithium-Ion Batteries Using an Ensemble Learning Framework Based on Optimized Interval of Charging Curves. *Energy Fuels* **2026**, *40*, 9073–9084. [CrossRef]
7. Chen, J.; Liu, Y.; Yong, J.; Yang, C.; Yan, L.; Zheng, Y. State of Health Estimation of Lithium-Ion Batteries Using Fusion Health Indicator by PSO-ELM Model. *Batteries* **2024**, *10*, 380. [CrossRef]
8. Wang, Y.; Zhang, X.; Chen, D.; Yang, J. State of health estimation of lithium-ion batteries based on BKA-FSVR algorithm with feature reconstruction from partial constant current charging interval. *Energy* **2025**, *335*, 137888. [CrossRef]
9. Petkovski, E.; Marri, I.; Cristaldi, L.; Faifer, M. State of Health Estimation Procedure for Lithium-Ion Batteries Using Partial Discharge Data and Support Vector Regression. *Energies* **2024**, *17*, 206. [CrossRef]
10. Liu, Z.; Zhao, J.; Wang, H.; Yang, C. A New Lithium-Ion Battery SOH Estimation Method Based on an Indirect Enhanced Health Indicator and Support Vector Regression in PHMs. *Energies* **2020**, *13*, 830. [CrossRef]
11. Yu, J.; Mo, B.; Tang, D.; Yang, J.; Wan, J.; Liu, J. Indirect State-of-Health Estimation for Lithium-Ion Batteries under Randomized Use. *Energies* **2017**, *10*, 2012. [CrossRef]
12. Chen, B.; Li, S.; Zhou, N.; Li, Y.; Zhang, Q. Physics-Consistent Domain-Aware SOH Estimation for Cross-Cell Battery Health Prediction. *Batteries* **2026**, *12*, 340. [CrossRef]
13. Yang, H.; Li, J.; Cen, K.; Yan, J.; Bo, Z. A cross-cell voltage interval mining and pyramid attention based framework for state-of-health estimation under incomplete data. *J. Power Sources* **2026**, *663*, 238918. [CrossRef]
14. Barua, N.; Collins, M.D.; Islam, M.S.; Barua, A.; Siddiquee, K.N.A. Predictor Structure Modulates Validation Inflation and Cell-Transfer Reliability in Battery State-of-Health Estimation. *Batteries* **2026**, *12*, 342. [CrossRef]
15. Le, H.H.; Nguyen, K.-A. Charging phase health indicators for battery state-of-health estimation: A systematic comparison of CC, CV, and combined approaches under cross-battery validation. *Eksploat. Niezawodn. Maint. Reliab.* **2026**, *28*, 220211. [CrossRef]
16. Zhuo, S.; Zou, F.; Liao, L.; Cai, X. MI-SOH: A multi-indicator feature dependency model for lithium-ion battery state-of-health estimation. *Sci. Rep.* **2026**, *16*, 12309. [CrossRef] [PubMed]
17. Saha, B.; Goebel, K. *Battery Data Set*; NASA Prognostics Data Repository, NASA Ames Research Center: Moffett Field, CA, USA, 2007.
18. Center for Advanced Life Cycle Engineering (CALCE), University of Maryland. Battery Research Data: CS2 Battery. Available online: https://calce.umd.edu/battery-data (accessed on 21 September 2026).
19. He, W.; Williard, N.; Osterman, M.; Pecht, M. Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method. *J. Power Sources* **2011**, *196*, 10314–10321. [CrossRef]
20. Williard, N.; He, W.; Osterman, M.; Pecht, M. Comparative Analysis of Features for Determining State of Health in Lithium-Ion Batteries. *Int. J. Progn. Health Manag.* **2013**, *4*. [CrossRef]
21. Xing, Y.; Ma, E.W.M.; Tsui, K.-L.; Pecht, M. An ensemble model for predicting the remaining useful performance of lithium-ion batteries. *Microelectron. Reliab.* **2013**, *53*, 811–820. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.