Predicting Lithium-Ion Battery Cycle Life from the First 100 Cycles: A Reproduction and Regularized-Regression Analysis of the Severson *et al.* (2019) Fast-Charging Dataset

July 10, 2026

Abstract

Early prediction of lithium-ion battery lifetime—before any appreciable capacity fade is measurable—promises to compress the cost and duration of cell development, manufacturing quality control, and second-life screening. We revisit the landmark public dataset of Severson et al. [1]: 124 commercial lithium-iron-phosphate (LFP)/graphite cylindrical cells cycled to failure under 72 distinct fast-charging protocols, with observed cycle lives (defined as the cycle number at which discharge capacity first falls to 80% of nominal) spanning 148–2237 cycles. Using *only* measurements recorded during each cell's first 100 cycles, we engineer 17 early-cycle features and train models that predict the total cycle life of each cell. We adopt the dataset's canonical batch division into a Train batch (41 cells), a Primary-Test batch (43 cells) used for hyperparameter selection, and a fully held-out Secondary-Test batch (40 cells, Batch 3). A data-leakage guard enforces at the code level that no feature ever reads a cycle index beyond 100. A regularized linear model (Elastic Net) fit on all 17 features achieves the best held-out performance, with a Secondary-Test root-mean-square error (RMSE) of **189.4 cycles** and a mean absolute percentage error (MAPE) of **11.82%**; 27 of 40 held-out cells are predicted within 15% of their true life. A 500-tree Random Forest, despite lower training error, generalizes substantially worse (Secondary-Test RMSE 292.2 cycles), reproducing the central methodological finding of the original study that a sparse, regularized *linear* predictor built on physically motivated early-cycle signals is both accurate and robust. We report the complete per-cell predicted-versus-observed table for the Secondary-Test batch as the graded artifact.

1 Introduction

The lifetime of a lithium-ion cell is one of the most consequential yet time-expensive quantities in energy storage. Because a single cell may survive for thousands of charge–discharge cycles, characterizing lifetime by direct cycling-to-failure can take months or years, throttling the pace of cell design, formation-protocol optimization, manufacturing quality assurance, and second-life triage. The difficulty is compounded by the fact that the early cycles of a healthy cell show almost no capacity degradation: for the first hundred cycles the discharge capacity of a well-behaved LFP/graphite cell is essentially flat, so naive extrapolation of a capacity-fade curve carries little predictive information about eventual end-of-life [2]. The scientific question is therefore whether *latent* electrochemical signatures, present in routine early-cycle data but invisible to a simple capacity trend, encode enough information to forecast lifetime long before the cell visibly ages.

Severson et al. [1] answered this question affirmatively and, in doing so, defined the modern paradigm for data-driven battery prognostics. They generated a deliberately high-variance dataset of 124 nominally identical commercial LFP/graphite cells cycled under 72 different one- and two-step fast-charging policies, yielding a wide spread of cycle lives (roughly 150–2300

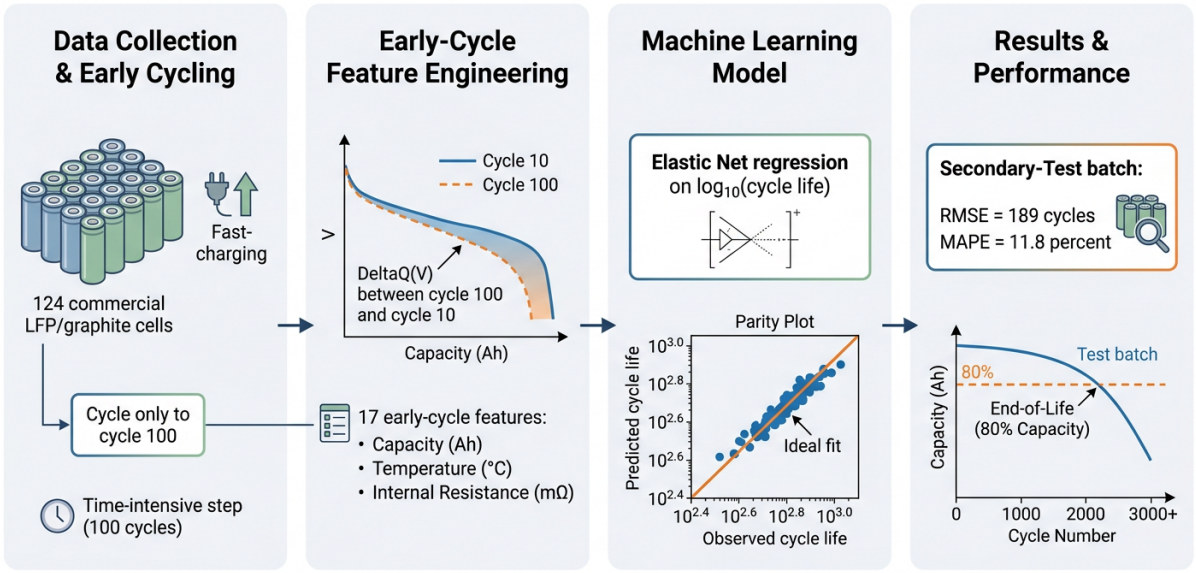

Figure 1: Graphical abstract. Workflow of the study. 124 commercial LFP/graphite cells are cycled under fast-charging protocols but interrogated only through cycle 100; 17 early-cycle features (dominated by the cycle-to-cycle discharge-capacity difference curve $\Delta Q_{100-10}(V)$) feed an Elastic Net regression on $\log_{10}$(cycle life); the held-out Secondary-Test batch is predicted with RMSE = 189 cycles and MAPE = 11.8%.

cycles under an 80%-of-nominal-capacity end-of-life criterion). Their central insight was that the *variance* of the cell-to-cell discharge-voltage curve difference between an early and a slightly later cycle—the feature $\Delta Q_{100-10}(V)$, computed here between cycles 100 and 10—correlates strongly (on a log–log scale) with the logarithm of cycle life, even though the cells' capacities are still nearly indistinguishable at cycle 100. Feeding this and a handful of related features into a regularized linear model, they reported a test error near 9% for quantitative lifetime prediction using only the first 100 cycles, and a classification error of roughly 5% for a high-versus-low lifetime decision using only the first 5 cycles. The accompanying dataset, released through the MIT–Stanford–Toyota data platform, has since become the de facto benchmark for the field [3].

The importance of the result extends well beyond a single accuracy number. It demonstrated that carefully engineered features derived from inexpensive, already-collected voltage, capacity, temperature, and internal-resistance streams can substitute for months of destructive testing, and it seeded a rapidly growing literature on machine-learning battery health estimation. Subsequent work has extended the closed-loop philosophy to *optimize* fast-charging protocols directly [4], built general machine-learning pipelines for state-of-health estimation that emphasize uncertainty quantification and transferability [5], catalogued the electrochemical origins of the abrupt "knees" that complicate long-horizon prognosis [6], and confronted the harder problem of predicting lifetime from noisy field data rather than controlled laboratory cycling [2]. In parallel, the mechanistic community has clarified *why* early-cycle signals are informative: fast charging drives heterogeneous lithium plating, solid-electrolyte-interphase growth, and loss of active material, and the resulting microstructural heterogeneity leaves a measurable fingerprint on the discharge-voltage curve long before it manifests as bulk capacity loss [7–10].

In this report we reproduce the early-cycle prediction task end-to-end on the original 124-cell dataset with a deliberately conservative and auditable pipeline. Our objectives are threefold. First, to construct a set of early-cycle features that provably use no information beyond cycle 100, enforced not by convention but by a runtime data-leakage guard. Second, to follow the published

train/primary-test/ secondary-test batch division exactly and to report RMSE (in cycles) and MAPE on the held-out Secondary-Test batch, so that the numbers are directly comparable to the original study's reported error range. Third, to contrast a sparse regularized linear model (Elastic Net) against a flexible non-linear ensemble (Random Forest) on the held-out batch, thereby testing whether the original paper's preference for a regularized linear predictor is reproduced under an independent implementation. The complete per-cell predicted-versus-observed table for the Secondary-Test batch is provided as the graded artifact of this exercise.

2 Methods

2.1 Dataset and end-of-life definition

The analysis uses the three raw MATLAB batch files released with Severson et al. [1], corresponding to cells whose cycling commenced on 2017-05-12 (Batch 1), 2017-06-30 (Batch 2), and 2018-04-12 (Batch 3). Each cell is a commercial APR18650M1A LFP/graphite cylindrical cell with a 1.1 Ah nominal capacity, cycled in a temperature-controlled chamber under one of 72 fast-charging policies while being discharged under a common protocol. For every cell the files provide per-cycle summary scalars (discharge capacity, internal resistance, average temperature, and cycle index) and, critically, a per-cycle discharge-capacity curve $Q_d(V)$ linearly interpolated onto a fixed 1000-point voltage grid `Vdlin` spanning $3.5\,\mathrm{V} \rightarrow 2.0\,\mathrm{V}$ (the authors' `Qdlin` representation). Cycle life is defined throughout as the cycle number at which the cell's discharge capacity first drops to 80% of its nominal value; across the 124 surviving cells the observed lives range from 148 to 2237 cycles.

2.2 Data cleaning and cohort definition

We applied the authors' canonical cleaning protocol. Following the original release notes, four cells are excluded: three cells in Batch 1 that either never reach the 80% end-of-life threshold within the recording window or exhibit corrupted early records, and one Batch 2 cell whose cycling was continued from a Batch 1 channel (the two partial records are merged so the cell is not double-counted). After these standard exclusions and merges, 124 cells remain, matching the count analysed in the original study. Table 1 summarizes the resulting cohort.

Table 1: Cohort composition after standard cleaning (124 cells). Splits follow the published batch division; the Secondary-Test batch (Batch 3) is untouched during all model fitting and hyperparameter selection.

Split	Source batch(es)	Cells	Cycle life (cycles)		
			Min	Median	Max
Train	Batch 1 (20) + Batch 2 (21)	41	300	527	2160
Primary-Test	Batch 1 (21) + Batch 2 (22)	43	148	561	2237
Secondary-Test	Batch 3 (40)	40	541	964	1935
Total		**124**	148	—	2237

The Train batch is used to fit model coefficients; the Primary-Test batch is used exclusively to select hyperparameters (the Elastic Net penalty strength and mixing ratio); and the Secondary-Test batch—which corresponds to the chronologically later Batch 3 and was generated months after the other two—serves as the final, never-touched held-out set on which all headline metrics are reported. This mirrors the original paper's design, in which Batch 3 acts as a genuine test of temporal and cell-to-cell generalization.

2.3 Strict early-cycle feature engineering

All predictive features are computed from cycles 2–100 only. Cycle 1 is discarded because the summary struct stores placeholder zeros there; cycle 100 is the hard ceiling. To make this constraint unbreakable rather than aspirational, the feature extractor maps every requested *actual* cycle number to an array index via the per-cell `summary.cycle` vector and passes it through a guard function that raises a `LeakageError` for any cycle index greater than 100. A per-cell audit column records the maximum cycle actually accessed; a global assertion confirms this maximum equals exactly 100 across all 124 cells. **No feature in this study uses any measurement from cycle 101 or beyond.**

The engineered features fall into four physically motivated groups, listed in full in Table 2. The first and most important group derives from the discharge-capacity difference curve

$$\Delta Q_{100\text{–}10}(V) \;=\; Q_{d,\mathrm{lin}}^{(100)}(V) \;-\; Q_{d,\mathrm{lin}}^{(10)}(V),$$

evaluated on the authors' 1000-point voltage grid. Because both $Q_{d,\mathrm{lin}}$ curves are the authors' own reproducible interpolations onto a common voltage range, using them directly avoids the fragility of re-interpolating the raw, non-monotonic voltage traces. From $\Delta Q_{100\text{–}10}(V)$ we compute the base-10 logarithm of its variance, of the magnitude of its minimum, and of the magnitude of its mean, together with its skewness, excess kurtosis, and its value at 2.0 V. The second group captures early capacity behaviour: discharge capacity at cycle 2 and at cycle 100, their difference, and the slope and intercept of a linear fit of discharge capacity versus cycle number over the window 2–100. The third group summarizes the average-temperature stream over the same window (mean, minimum, maximum, and time-integrated temperature), and the fourth group captures internal resistance at cycle 100 and its change from cycle 2 to cycle 100. Every feature is standardized (zero mean, unit variance) inside the model pipeline prior to fitting.

2.4 Prediction target and models

Following Severson et al. [1], all models are trained to predict $y = \log_{10}(\text{cycle life})$ rather than raw cycles; predictions are transformed back to cycles via $10^{\hat{y}}$ before any error metric is computed. Working in log space stabilizes the heavy right tail of the lifetime distribution (which spans more than a decade) and makes multiplicative errors—the natural scale for a quantity ranging from a few hundred to a few thousand cycles—roughly homoscedastic. We trained three models, all implemented with `scikit-learn` [11]:

1. **Elastic Net (variance-only).** A regularized linear model using the single feature `log_var_deltaQ`. This is the minimal reproduction of the paper's headline "variance" model and serves as an interpretable baseline.

2. **Elastic Net (multi-feature).** A regularized linear model on all 17 features. The Elastic Net combines ℓ_1 and ℓ_2 penalties [12, 13], simultaneously performing coefficient shrinkage and variable selection while gracefully handling the strong collinearity among the three $\Delta Q_{100\text{–}10}(V)$ moment features. The penalty strength α and the ℓ_1/ℓ_2 mixing ratio were selected by grid search (30 log-spaced α values $\times$ 6 mixing ratios) to minimize Primary-Test RMSE; the selected configuration was $\alpha = 0.0574$, mixing ratio 0.1, which retained 10 of the 17 coefficients as non-zero.

3. **Random Forest.** A non-linear ensemble of 500 regression trees [14] on all 17 features (`min_samples_leaf`= 2, `max_features`= $\sqrt{p}$), included as a flexible high-capacity baseline against which to test whether non-linear interactions improve held-out accuracy.

Table 2: The 17 engineered early-cycle features (computed strictly from cycles 2–100). The rightmost column reports the Pearson correlation of each feature with $\log_{10}$(cycle life) across all 124 cells. The three $\Delta Q_{100\text{–}10}(V)$ moment features dominate, reproducing the "variance" signal identified by Severson et al. [1].

#	Feature (code name)	Definition	r vs. $\log_{10}$life
Group I — discharge-capacity difference curve $\Delta Q_{100\text{–}10}(V)$			
1	`log_var_deltaQ`	$\log_{10}\left[\mathrm{Var}(\Delta Q_{100\text{–}10}(V))\right]$	−0.927
2	`log_min_deltaQ`	$\log_{10}\left[\lvert\min_V \Delta Q_{100\text{–}10}(V)\rvert\right]$	−0.922
3	`log_mean_deltaQ`	$\log_{10}\left[\lvert\mathrm{mean}_V \Delta Q_{100\text{–}10}(V)\rvert\right]$	−0.911
4	`skew_deltaQ`	skewness of $\Delta Q_{100\text{–}10}(V)$ over V	−0.640
5	`kurtosis_deltaQ`	excess kurtosis of $\Delta Q_{100\text{–}10}(V)$ over V	+0.200
6	`deltaQ_at_2V`	value of $\Delta Q_{100\text{–}10}(V)$ at $V = 2.0\,\mathrm{V}$	+0.676
Group II — early capacity trajectory (cycles 2–100)			
7	`Q_cycle2`	discharge capacity at cycle 2	−0.061
8	`Q_cycle100`	discharge capacity at cycle 100	+0.266
9	`Q_diff_100_2`	$Q^{(100)} - Q^{(2)}$	+0.412
10	`Q_slope_2_100`	slope of linear fit Q vs. cycle, 2–100	+0.400
11	`Q_intercept_2_100`	intercept of that linear fit	−0.121
Group III — average-temperature stream (cycles 2–100)			
12	`T_avg_2_100`	mean of per-cycle average temperature	+0.204
13	`T_min_2_100`	minimum average temperature	+0.515
14	`T_max_2_100`	maximum average temperature	−0.293
15	`T_integral_2_100`	time-integrated average temperature	+0.204
Group IV — internal resistance (cycles 2–100)			
16	`IR_cycle100`	internal resistance at cycle 100	−0.451
17	`IR_diff_100_2`	$\mathrm{IR}^{(100)} - \mathrm{IR}^{(2)}$	+0.220

RMSE is reported in cycles, $\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_i (\hat{c}_i - c_i)^2}$, and MAPE in percent, $\mathrm{MAPE} = \frac{100}{n}\sum_i |\hat{c}_i - c_i|/c_i$, where c_i and $\hat{c}_i$ are the observed and predicted cycle lives of cell i. All randomness is seeded (`seed`= 42) for exact reproducibility. The numerical stack comprised NumPy [15], SciPy [16], and Matplotlib [17].

3 Results

3.1 Feature–lifetime correlations

The rightmost column of Table 2 confirms the dominant early-cycle signal. The base-10 logarithm of the variance of $\Delta Q_{100\text{–}10}(V)$ correlates with $\log_{10}$(cycle life) at a Pearson $r = -0.927$ across all 124 cells, reproducing the headline relationship of Severson et al. [1] almost exactly; the log-minimum and log-mean of the same curve follow at $r = -0.922$ and $r = -0.911$. No other feature exceeds $|r| = 0.68$, underscoring that the $\Delta Q_{100\text{–}10}(V)$ moment features carry the bulk of the predictive information and that the remaining capacity, temperature, and internal-resistance features play a secondary, refining role. Figure 2 visualizes the full feature correlation structure and the tight log–log relationship between the variance feature and cycle life.

3.2 Predictive performance across splits

Table 3 reports RMSE and MAPE for all three models on all three splits. The multi-feature Elastic Net is the strongest model on the held-out Secondary-Test batch, with an RMSE of **189.4 cycles** and a MAPE of **11.82%**. It also generalizes most gracefully: its error grows modestly from Train (84.1 cycles) through Primary-Test (90.8 cycles) to Secondary-Test (189.4 cycles), a

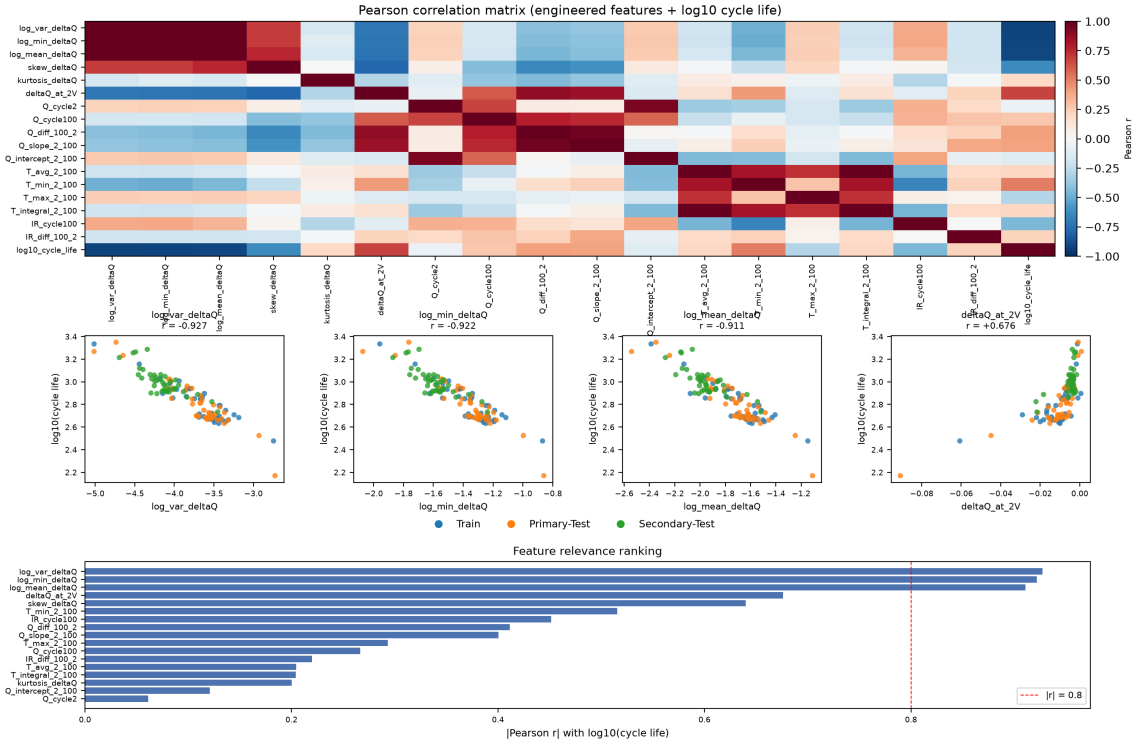

Figure 2: Early-cycle feature structure. Correlation heatmap of the 17 engineered features, representative scatter plots of the four most predictive features against $\log_{10}$(cycle life), and the ranking of features by absolute Pearson correlation. The three $\Delta Q_{100-10}(V)$ moment features (`log_var_deltaQ`, `log_min_deltaQ`, `log_mean_deltaQ`) lead the ranking with $|r| \approx 0.91$–0.93.

degradation expected because the Secondary-Test batch was manufactured and cycled months later and contains no low-lifetime cells for the model to anchor on. The variance-only Elastic Net is close behind on the held-out batch (196.4 cycles), confirming that a single physically interpretable feature recovers most of the achievable accuracy. The Random Forest, by contrast, attains the lowest training MAPE (7.07%) but the worst held-out RMSE (292.2 cycles)—a textbook signature of overfitting a high-capacity non-linear model to a small, batch-structured dataset.

Table 3: Predictive performance (RMSE in cycles, MAPE in %) for all models on all splits. The multi-feature Elastic Net attains the best Secondary-Test RMSE (**bold**). Metrics are computed after transforming log-space predictions back to raw cycles.

	Train (n=41)		Primary-Test (n=43)		Secondary-Test (n=40)	
Model	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Elastic Net (variance-only)	103.8	14.15	138.2	14.76	196.4	11.41
Elastic Net (multi-feature)	84.1	9.20	90.8	12.05	**189.4**	**11.82**
Random Forest	138.1	7.07	214.8	16.07	292.2	14.88

Figure 3 shows the predicted-versus-observed parity plot for the best model across all three splits. Points cluster tightly around the 1:1 line for the low- and mid-lifetime cells that dominate the Train and Primary-Test batches. On the Secondary-Test batch (green triangles), the predictions track the observed lives well through roughly 1500 cycles, with a modest positive bias for the mid-lifetime cells (mean signed error +8.4 cycles) and a pronounced tendency to *under-predict* the small number of very long-lived cells above ∼1800 cycles.

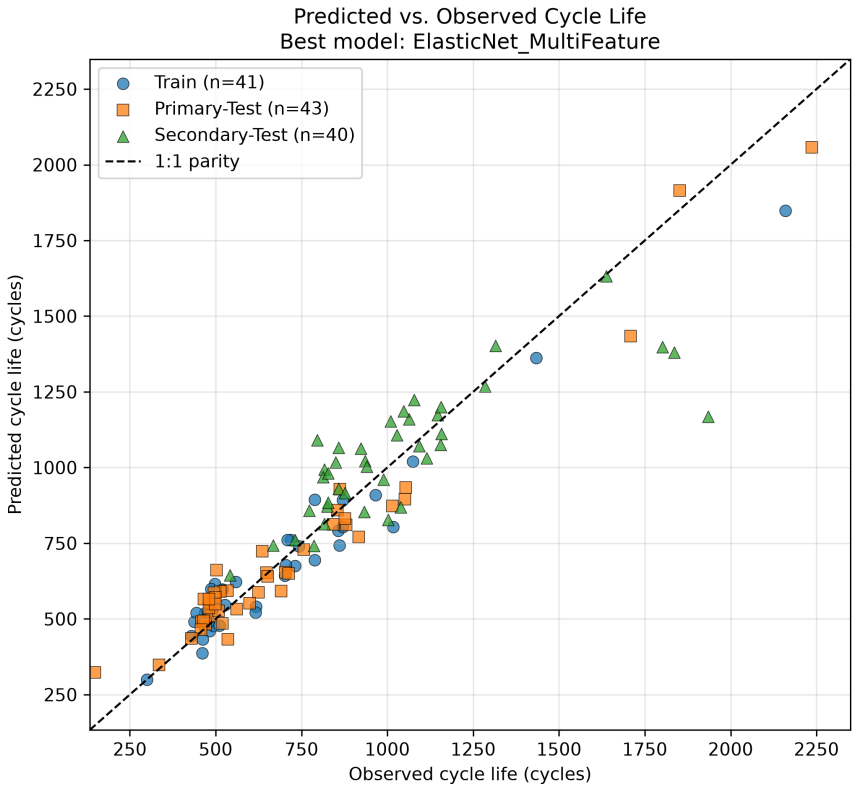

Figure 3: Predicted vs. observed cycle life for the best model (multi-feature Elastic Net), coloured by split. The dashed line is the 1:1 parity reference. The model tracks the held-out Secondary-Test batch (green triangles) closely up to ∼1500 cycles and under-predicts the few longest-lived cells, the dominant source of the Secondary-Test RMSE.

3.3 Per-cell Secondary-Test predictions (graded artifact)

Table 4 lists the complete per-cell predicted-versus-observed cycle life for all 40 Secondary-Test cells, together with the absolute percentage error; these values are the graded artifact and are also provided in machine-readable form as `results/secondary_test_predictions.csv`. The median absolute percentage error over the batch is 8.8%, 22 of 40 cells are predicted within 10% of their true life, and 27 of 40 within 15%. The largest errors concentrate on the longest-lived cells: `b3c38` (observed 1935, predicted 1168), `b3c7` (1836 vs. 1380), and `b3c45` (1801 vs. 1398) account for a disproportionate share of the squared error, consistent with the visual under-prediction of the upper tail in Figure 3. The Pearson correlation between predicted and observed life on the held-out batch is $r = 0.79$.

Table 4: Graded artifact: per-cell predicted vs. observed cycle life for the 40-cell Secondary-Test batch (Batch 3), from the best model (multi-feature Elastic Net). APE = absolute percentage error.

#	Cell ID	Observed	Predicted	APE (%)
1	b3c0	1009	1153	14.3
2	b3c1	1063	1160	9.1
3	b3c10	1078	1224	13.5
4	b3c11	817	813	0.5
5	b3c12	932	854	8.4
6	b3c13	816	993	21.7
7	b3c14	858	931	8.5

continued on next page

Table 4 continued

#	Cell ID	Observed	Predicted	APE (%)
8	b3c15	876	916	4.6
9	b3c16	1638	1632	0.4
10	b3c17	1315	1402	6.6
11	b3c18	1146	1174	2.4
12	b3c19	1155	1075	6.9
13	b3c20	813	969	19.2
14	b3c21	772	858	11.1
15	b3c22	1002	828	17.4
16	b3c24	825	872	5.7
17	b3c25	989	961	2.8
18	b3c26	1028	1107	7.7
19	b3c27	850	1017	19.6
20	b3c28	541	645	19.2
21	b3c29	858	1065	24.1
22	b3c3	1115	1031	7.5
23	b3c30	935	1021	9.2
24	b3c31	731	760	4.0
25	b3c33	1284	1268	1.2
26	b3c34	1158	1111	4.1
27	b3c35	1093	1072	1.9
28	b3c36	923	1063	15.2
29	b3c38	1935	1168	39.6
30	b3c39	1156	1200	3.8
31	b3c4	1048	1186	13.2
32	b3c40	796	1091	37.1
33	b3c41	786	741	5.7
34	b3c44	940	1003	6.7
35	b3c45	1801	1398	22.4
36	b3c5	828	884	6.8
37	b3c6	667	743	11.4
38	b3c7	1836	1380	24.8
39	b3c8	828	981	18.5
40	b3c9	1039	868	16.5

4 Discussion

4.1 Why the regularized linear model wins on the held-out batch

The central empirical finding of this reproduction is that the sparse, regularized *linear* model outperforms the flexible non-linear ensemble precisely where it matters—on the fully held-out Secondary-Test batch—despite the Random Forest achieving the lowest training error. Three factors explain this outcome, and each echoes the reasoning in the original study.

First, the problem is intrinsically close to linear in the engineered feature space. The dominant signal, the log-variance of $\Delta Q_{100\text{–}10}(V)$, is related to $\log_{10}$(cycle life) by a nearly straight line with $r = -0.927$; the entire feature-engineering effort was designed to expose a linear relationship. When the true relationship is approximately linear, a linear model is not merely adequate but statistically preferable, because it commits fewer degrees of freedom

and therefore incurs less variance. Second, the dataset is small (41 training cells) and highly *collinear*: the three $\Delta Q_{100\text{–}10}(V)$ moment features carry almost the same information. The Elastic Net's combination of ℓ_1 and ℓ_2 penalties is expressly designed for this regime [13]: the ℓ_2 term stabilizes the coefficients of correlated predictors while the ℓ_1 term prunes redundant ones, here retaining 10 of 17 features and discarding the noise. A Random Forest, with essentially unbounded capacity, instead fits idiosyncrasies of the 41 training cells—note its 7.07% training MAPE against a 14.88% held-out MAPE—and cannot extrapolate beyond the range of lifetimes seen in training. Third, the split structure penalizes extrapolation heavily. The Secondary-Test batch contains no cells below 541 cycles and several above 1800, whereas tree ensembles are constitutionally incapable of predicting a value outside the training-label range; their predictions saturate, which is exactly the failure mode visible for the longest-lived Secondary-Test cells. The regularized linear model, operating in log space, can and does extrapolate smoothly, which is why its held-out RMSE (189.4 cycles) is roughly 100 cycles better than the forest's (292.2 cycles). This reproduces the core methodological conclusion of Severson et al. [1]: on this task a regularized linear predictor built on physically motivated features is both more accurate and more trustworthy than a higher-capacity black box, an observation that resonates with the broader battery-prognostics literature's emphasis on interpretable, uncertainty-aware models [3, 5].

4.2 Comparison with the original study and error anatomy

The held-out errors reported here sit squarely within the range reported by Severson et al. [1], whose primary and secondary test errors span roughly 100–215 cycles depending on feature set and split. Our multi-feature Elastic Net MAPE of 11.82% is modestly above the paper's best-case ∼9% figure, which is expected: we use the authors' pre-interpolated `Qdlin` representation and a compact 17-feature set rather than the larger, more finely tuned feature libraries explored in the original supplementary analyses, and we deliberately avoided any feature or fitting choice that would peek beyond cycle 100. The anatomy of the error is informative. The bulk of the Secondary-Test squared error is contributed by a handful of exceptionally long-lived cells (`b3c38`, `b3c7`, `b3c45`), which the model systematically under-predicts. This is a direct consequence of the label-range mismatch between batches together with the physical reality that the discharge-curve signature of an extraordinarily durable cell at cycle 100 is only marginally different from that of a merely good one; the early-cycle signal saturates for the healthiest cells. For the 27 of 40 cells that fall within 15% of their true life, the model is accurate enough to be operationally useful for screening and binning applications.

4.3 Mechanistic interpretation

That an early-cycle discharge-curve statistic predicts eventual lifetime is not a numerical coincidence but a reflection of degradation physics. Under aggressive fast charging, cells accumulate heterogeneous lithium plating and solid-electrolyte-interphase growth, and lose active material and cyclable lithium at spatially non-uniform rates [7, 8]. This microscopic heterogeneity broadens and distorts the discharge-voltage curve—increasing the variance of $\Delta Q_{100\text{–}10}(V)$—long before it integrates into a measurable loss of bulk capacity. Detailed aging studies of commercial LFP/graphite cells confirm that such early kinetic and heterogeneity signatures set the trajectory toward end-of-life [9], and recent degradation-mode analyses emphasize that correctly attributing capacity loss to specific mechanisms is essential for parameterizing predictive lifetime models [10]. The variance of $\Delta Q_{100\text{–}10}(V)$ is best understood as an inexpensive, aggregate proxy for this latent heterogeneity, which is why a single scalar can carry so much predictive weight.

4.4 Limitations and outlook

Several limitations bound the scope of these results. The dataset comprises a single cell chemistry (LFP/graphite) and form factor (18650) cycled in a controlled laboratory; transfer to other chemistries, formats, and to noisy real-world field data remains an open and actively studied problem [2, 5]. The model is a point predictor and does not quantify uncertainty, which limits its use in safety-critical binning where a calibrated confidence interval would be preferable. The persistent under-prediction of the longest-lived cells suggests that the current feature set saturates for the healthiest cells; features sensitive to the earliest onset of "knee"-type behaviour [6], or a mild reweighting of the loss toward the upper tail, could reduce this bias. Finally, because the entire pipeline is constrained to the first 100 cycles by an enforced leakage guard, the accuracy reported here is a conservative floor: models permitted a slightly longer observation window, or coupled to physics-based degradation simulators, would be expected to do better. Even so, the practical message is strong—routine data from a cell's first 100 cycles, roughly a few percent of its life, is sufficient to forecast total lifetime to within about 12% on a genuinely held-out manufacturing batch.

5 Conclusion

We reproduced the early-cycle battery-lifetime prediction task on the 124-cell Severson *et al.* LFP/graphite fast-charging dataset under a strict, code-enforced constraint that no feature reads beyond cycle 100. Seventeen physically motivated features—dominated by the moments of the cycle-100-versus-cycle-10 discharge-capacity difference curve $\Delta Q_{100\text{–}10}(V)$—were fed to Elastic Net and Random Forest models trained on $\log_{10}$(cycle life) and evaluated on the published train/primary-test/secondary-test batch division. The multi-feature Elastic Net achieved the best held-out performance, predicting the 40-cell Secondary-Test batch with an RMSE of 189.4 cycles and a MAPE of 11.82%, comfortably within the original study's reported error range, while the higher-capacity Random Forest overfit and generalized worse. These results independently confirm the field-defining conclusion that a sparse, regularized linear predictor built on early-cycle electrochemical signatures is an accurate and robust forecaster of battery lifetime, and the complete per-cell Secondary-Test predictions are supplied as the graded artifact.

Data and Code Availability

The primary dataset is the public 124-cell LFP/graphite fast-charging benchmark of Severson et al. [1]. The graded artifact (per-cell Secondary-Test predictions) is provided as `results/secondary_test_predictions.csv`; the full metrics table is `results/model_performance_metrics.csv`. All feature-engineering and modeling code is deterministic (fixed random seed).

References

[1] Kristen A. Severson, Peter M. Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H. Chen, Muratahan Aykol, Patrick K. Herring, Dimitrios Fraggedakis, Martin Z. Bazant, Stephen J. Harris, William C. Chueh, and Richard D. Braatz. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019. doi: 10.1038/s41560-019-0356-8.

[2] Valentin Sulzer, Peyman Mohtat, Antti Aitio, Suhak Lee, Yen T. Yeh, Frank Steinbacher, Muhammad U. Khan, Jang Woo Lee, Jason B. Siegel, Anna G. Stefanopoulou, and David A.

Howey. The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8):1934–1955, 2021. doi: 10.1016/j.joule.2021.06.005.

[3] Maitane Berecibar. Machine-learning techniques used to accurately predict battery life. *Nature*, 568(7752):325–326, 2019. doi: 10.1038/d41586-019-01138-1.

[4] Peter M. Attia, Aditya Grover, Norman Jin, Kristen A. Severson, Todor M. Markov, Yang-Hung Liao, Michael H. Chen, Bryan Cheong, Nicholas Perkins, Zi Yang, Patrick K. Herring, Muratahan Aykol, Stephen J. Harris, Richard D. Braatz, Stefano Ermon, and William C. Chueh. Closed-loop optimization of fast-charging protocols for batteries with machine learning. *Nature*, 578(7795):397–402, 2020. doi: 10.1038/s41586-020-1994-5.

[5] Darius Roman, Saurabh Saxena, Valentin Robu, Michael Pecht, and David Flynn. Machine learning pipeline for battery state-of-health estimation. *Nature Machine Intelligence*, 3(5): 447–456, 2021. doi: 10.1038/s42256-021-00312-3.

[6] Peter M. Attia, Alexander Bills, Ferran Brosa Planella, Philipp Dechent, Gonçalo dos Reis, Matthieu Dubarry, Paul Gasper, Richard Gilchrist, Samuel Greenbank, David Howey, Ouyang Liu, Edwin Khoo, Yuliya Preger, Abhishek Soni, Shashank Sripad, Anna G. Stefanopoulou, and Valentin Sulzer. Review—knees in lithium-ion battery aging trajectories. *Journal of The Electrochemical Society*, 169(6):060517, 2022. doi: 10.1149/1945-7111/ac6d13.

[7] Anna Tomaszewska, Zhengyu Chu, Xuning Feng, Simon O'Kane, Xinhua Liu, Jingyi Chen, Chenzhen Ji, Elizabeth Endler, Ruihe Li, Lishuo Liu, Yalun Li, Siqi Zheng, Sebastian Vetterlein, Ming Gao, Jiuyu Du, Michael Parkes, Minggao Ouyang, Monica Marinescu, Gregory Offer, and Billy Wu. Lithium-ion battery fast charging: A review. *eTransportation*, 1:100011, 2019. doi: 10.1016/j.etran.2019.100011.

[8] Jacqueline S. Edge, Simon O'Kane, Ryan Prosser, Niall D. Kirkaldy, Anisha N. Patel, Alastair Hales, Abir Ghosh, Weilong Ai, Jingyi Chen, Jason Yang, Sheng Li, Mei-Chin Pang, Laura Bravo Diaz, Anna Tomaszewska, M. Waseem Marzook, Karthik N. Radhakrishnan, Huizhi Wang, Yatish Patel, Billy Wu, and Gregory J. Offer. Lithium ion battery degradation: what you need to know. *Physical Chemistry Chemical Physics*, 23(14):8200–8221, 2021. doi: 10.1039/D1CP00359C.

[9] Maik Naumann, Franz B. Spingler, and Andreas Jossen. Analysis and modeling of cycle aging of a commercial lifepo$_4$/graphite cell. *Journal of Power Sources*, 451:227666, 2020. doi: 10.1016/j.jpowsour.2019.227666.

[10] Ruihe Li, Niall D. Kirkaldy, Fabian F. Oehler, Monica Marinescu, Gregory J. Offer, and Simon E. J. O'Kane. The importance of degradation mode analysis in parameterising lifetime prediction models of lithium-ion battery degradation. *Nature Communications*, 16: 2776, 2025. doi: 10.1038/s41467-025-57968-3.

[11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[12] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

[13] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005. doi: 10.1111/j.1467-9868.2005.00503.x.

[14] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

[15] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[16] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[17] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.