Journal of Dynamics, Monitoring and Diagnostics, 2026, 5, 117-130
https://doi.org/10.37965/jdmd.2026.1320

ISTP
RESEARCH ARTICLE

Prefix Treatment and Bayesian Partial Pooling for Early-Stage Lifetime Prediction from Degradation Data

Chizhi Chris Zhang,[1,3] Xinhui Feng,[1,2] and Jian Jiang[1,3]
[1]Advanced Computer and Digital Technology Research Center, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, Jilin, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Chinese Academy of Sciences, Beijing 100846, China

(Received 08 April 2026; Revised 11 May 2026; Accepted 25 May 2026; Published online 27 May 2026)

Abstract: Early-stage lifetime prediction from degradation data depends on how the observed prefix is represented. This article formulates prefix treatment as a statistical aggregation problem. Consecutive short-window extrapolations are converted into pseudo-lifetimes and regularized with Bayesian partial pooling, allowing information sharing across comparable units while preserving unit-level differences. The method is tested on gallium arsenide (GaAs) laser degradation data under a fixed six-measurement budget, with complete trajectories defining full-history pseudo-lifetime references. Bayesian partial pooling yields the lowest mean absolute error, 530.9 h, compared with 550.0 h for the strongest direct early-feature baseline and 947.0 h for early global fitting. The same principle is externally evaluated on the Oxford Battery Degradation Dataset using observed 95% capacity-retention threshold lives from a fixed early-cycle prefix. Bayesian partial pooling again gives the lowest leave-one-out mean absolute error, 21.0 cycles, compared with 38.9 cycles for window-median aggregation and 79.2 cycles for the strongest direct-feature baseline. Ablation results show that local-window calibration and window-specific reliability weighting carry most of the battery-side gain. The results support Bayesian partial pooling as an interpretable uncertainty-aware prefix treatment for noisy early degradation trajectories.

Keywords: Bayesian partial pooling; degradation monitoring; early-stage lifetime prediction; GaAs laser; lithium-ion battery; prefix treatment

LIST OF PRINCIPAL SYMBOLS

Symbol	Definition
D_i	Complete trajectory for unit i
$P_{i,m}$	Observed prefix containing the first m measurements
$u_{i,j}$,$z_{i,j}$	Usage index and degradation or health measurement
c_i	Threshold or reference definition for unit i
$g(\cdot)$	Fixed pseudo-lifetime map used within a validation setting
$W_{i,k}$	k-th local window extracted from the observed prefix
K_i	Number of local windows available for unit i
$L_{i,k}$	Local pseudo-lifetime obtained from window k
L_i^{glob}	Global-prefix pseudo-lifetime estimate
w_k	Normalized weight assigned to local window k
λ	Prefix-weight decay parameter
$y_{i,k}$	Log local pseudo-lifetime
θ_i	Device-specific latent log-lifetime
μ_θ	Fleet-level mean of latent log-lifetimes
σ_w^2	Within-device variance among local log pseudo-lifetimes
τ^2	Between-device variance in the partial-pooling model
$\bar{y}_i$	Mean of local log pseudo-lifetimes for unit i
α_i	Partial-pooling data weight for unit i
T_i	Bayesian pooled lifetime prediction for unit i

Corresponding author: Chizhi Chris Zhang (e-mail: zhangchizhi@ciomp.ac.cn).

I. INTRODUCTION

Condition monitoring, diagnostics, and prognostics use time-evolving observations to assess system health. Remaining useful life (RUL) prediction is central to this setting because it informs inspection planning, maintenance scheduling, and operational risk control. Foundational reviews show that prognostic accuracy depends not only on the selected degradation or learning model but also on how the monitored signal is represented, updated, and interpreted [1–3]. Early-stage prediction is especially difficult because only a short initial trajectory segment is available, the degradation signal is weak, and small perturbations in the first observations can strongly affect extrapolated failure time. Recent prognostics work has examined the broader RUL problem [4], while uncertainty-aware monitoring studies in bearing prognosis and SCADA-based prediction illustrate the practical range of this challenge [5–7]. Semiconductor laser degradation provides a stringent test case: degradation tests are costly, degradation can be slow, and maintenance-relevant decisions may be needed before run-to-failure information is available.

Existing work addresses this difficulty through three related lines of research. Review and early-prediction studies show that limited labels, incomplete life histories, and missing run-to-failure trajectories constrain RUL estimation. This issue appears in physics-informed RUL reviews [8], adaptive and data-model-linked early RUL prediction [9,10], probabilistic prediction without lifetime labels [11], and limited-data lithium-ion prognosis [12]. Small-sample learning studies seek robustness through interpretable

multiscale representations, semi-supervised fusion, and signal-processing priors [13–15]. Probabilistic degradation models represent uncertainty and heterogeneity directly. Bayesian process and hierarchical degradation models quantify uncertainty and allow information sharing across units [16–18]. Bayesian network formulations extend probabilistic inference to dependent or multi-state systems [19–21]. Dynamic decision models address deteriorating systems with probabilistic dependencies [22], fleet-level transfer supports knowledge sharing across engineering assets [23], and small-sample Bayesian reliability or calibrated Bayesian regression provide complementary routes for sparse reliability prediction [24,25].

Degradation-specific studies are also relevant. Bayesian hierarchical battery prognostics and battery condition-monitoring models show the value of probabilistic health modeling in lithium-ion systems [26–28]. Heterogeneous degradation and laser-data models address device-to-device variability and Wiener-process degradation [29,30], dynamic degradation paths, and semiparametric Bayesian modeling for laser data [31,32]. Related degradation and RUL models further treat model uncertainty and random effects [33,34], together with aleatory or epistemic uncertainty and maintenance decisions for heterogeneous items [35,36]. These studies establish probabilistic modeling as a natural framework for sparse degradation data, but most begin after the early observation prefix has already been chosen and summarized. Here, the early-information budget and pseudo-lifetime map are fixed, and the aggregation and regularization of prefix-derived pseudo-lifetimes become the central methodological problem.

The contribution is a fixed-information early-stage lifetime-prediction protocol that separates prefix treatment from physical failure-time observation and model-form selection; a Bayesian partial-pooling formulation that regularizes volatile local pseudo-lifetimes without assigning heuristic priority to the earliest windows; and an evaluation of the same prefix-treatment principle on a controlled gallium arsenide (GaAs) laser benchmark and the public Oxford lithium-ion battery dataset. Direct early-feature baselines, uncertainty intervals, residual diagnostics, and sensitivity analyses are included to test whether the observed gains are specific to a single modeling choice.

The analysis proceeds from the formal early-stage prediction setting to the partial-pooling model and then to two validations: a controlled GaAs laser benchmark with a full-history pseudo-lifetime reference and an Oxford battery study with observed 95% capacity-retention threshold lives.

II. EARLY-STAGE PROGNOSTIC SETTING AND COMPETING ESTIMATORS

A. MONITORING SETTING AND FIXED INFORMATION BUDGET

For a monitored unit indexed by i, the degradation or health record is written as an ordered sequence of paired usage and response measurements:

$$D_i = [(u_{i,1}, z_{i,1}), \ldots, (u_{i,n_i}, z_{i,n_i})]$$

The usage coordinate u may be inspection time, cycle count, or another monotone exposure index; z is the corresponding degradation or health measurement. A lifetime endpoint is specified either by a physical threshold or, when failure is not observed, by a consistent reference construction.

Prediction is required after only the first m observations have been collected. This prefix is the sole input to the early-stage predictor; later observations are held back for evaluation through either an observed threshold life or a predefined full-history reference. The task is therefore constrained by monitoring information, not merely by a small sample size.

The framework applies to degradation datasets in which several comparable units are observed under broadly aligned conditions, a monotone scalar or low-dimensional health indicator is available, and a common threshold can be defined. It is most useful when short prefixes produce several informative but noisy local extrapolations and when exchangeability across units is scientifically defensible. It is less appropriate for fleets governed by distinct degradation mechanisms or for settings in which rich early covariates already yield a stable lifetime ranking without pseudo-lifetime aggregation.

B. COMMON PSEUDO-LIFETIME MAP

The pseudo-lifetime map is fixed before prefix treatments are compared so that the experiment tests how early information is summarized rather than which degradation law is selected. In the notation below, g is the fixed map, S is an admissible trajectory segment, c_i is the unit-specific threshold or reference definition, and η contains map settings or calibration constants. The generic map is written as:

$$L(S) = g(S; c_i, \eta) \tag{1}$$

For a global-prefix fit, the admissible segment S is the observed prefix $P_{i,m}$; substituting it into the same map gives

$$L_i^{glob} = g(P_{i,m}; c_i, \eta) \tag{2}$$

This restriction is methodological rather than physical. The main experiments do not search over alternative degradation laws or lifetime distributions. Instead, g is fixed within each validation setting, which attributes differences in prediction to the aggregation and pooling of the same prefix information.

The validation sections specify the threshold choices, reference construction, and map-adequacy checks used in each dataset.

C. LOCAL PSEUDO-LIFETIMES AND PREFIX INSTABILITY

The selected map can be applied either to the whole prefix or to short consecutive windows within it. The global representation fits the map directly to $P_{i,m}$. The local representation converts each consecutive pair or short window into a pseudo-lifetime. For a window length r, the kth local window is

$$W_{i,k} = [(u_{i,k}, z_{i,k}), \ldots, (u_{i,k+r-1}, z_{i,k+r-1})],\ k = 1, \ldots, K_i \tag{3}$$

Applying the same pseudo-lifetime map to this window gives

$$L_{i,k} = g(W_{i,k}; c_i, \eta),\ k = 1, \ldots, K_i \tag{4}$$

Collecting these values produces the local pseudo-lifetime vector

$$\ell_i = (L_{i,1}, \ldots, L_{i,K_i}) \tag{5}$$

These local pseudo-lifetimes are not independent observations. They quantify the instability of short-window extrapolation when only prefix information is available. When the window-level estimates are volatile, the subsequent weighting or pooling step can influence the final lifetime prediction as much as the raw degradation measurements. The fixed-prefix early-stage prediction protocol is summarized in Fig. 1, and the consecutive prefix windows used to obtain local pseudo-lifetime estimates are illustrated in Fig. 2.

D. PREFIX AGGREGATION AND EMPHASIS STRATEGIES

The weighting schemes do not assume that the earliest calendar times are intrinsically more reliable. They quantify how much the final prediction changes when extra weight is placed on the earliest local windows. With $L_{i,k}$ denoting the pseudo-lifetime from window k of unit i, the normalized prefix weights are

$$w_k \propto e^{-\lambda(k-1)},\ k = 1, \ldots, K_i, \sum_{k=1}^{K_i} w_k = 1 \tag{6}$$

where λ controls the degree of emphasis on the earliest windows. The weighted prediction is

$$T_{i(\lambda)} = \sum_{k=1}^{K_i} w_k L_{i,k} \tag{7}$$

The weighting rule is used as a sensitivity device, not as a physical law. It measures how strongly heuristic emphasis on the first local windows can alter the downstream lifetime prediction.

The nonuniform weight profiles in Fig. 3 serve as sensitivity baselines. The window median provides a robust nonparametric aggregation rule, and the two initial-emphasis estimators test the effect of assigning greater leverage to the earliest local slopes. Hard anomaly deletion is excluded from the main comparison because, with only a few windows, deleting an extreme local estimate can remove the instability that the early-stage protocol is intended to evaluate. The fixed-budget comparison therefore contains three estimator groups: the proposed Bayesian partial-pooling estimator, pseudo-lifetime comparators based on early global fitting, window-median aggregation, and two initial-emphasis rules, and direct-feature baselines extracted from the same early prefix. This grouping keeps the information budget fixed while separating alternatives inside and outside the proposed framework.

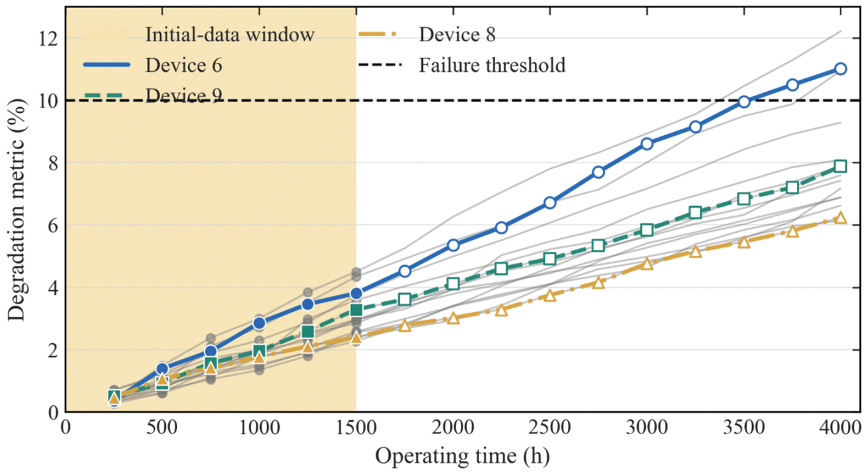

Fig. 1. Fixed-prefix early-stage lifetime prediction protocol.

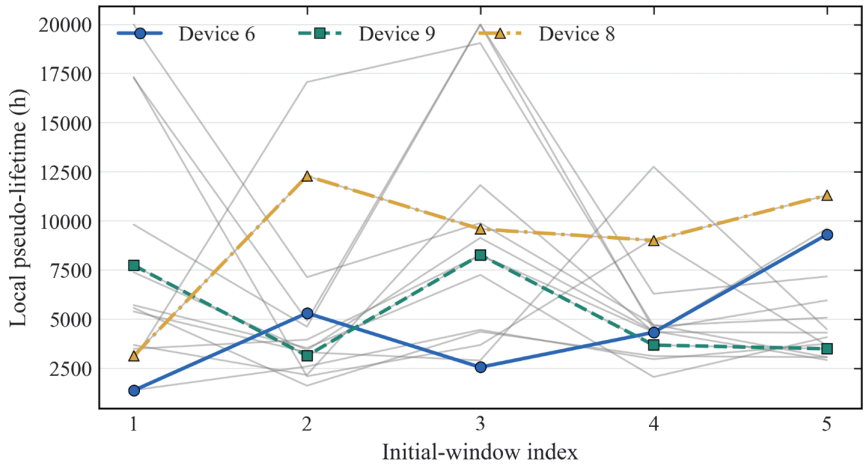

Fig. 2. Consecutive prefix windows and their local pseudo-lifetime estimates.

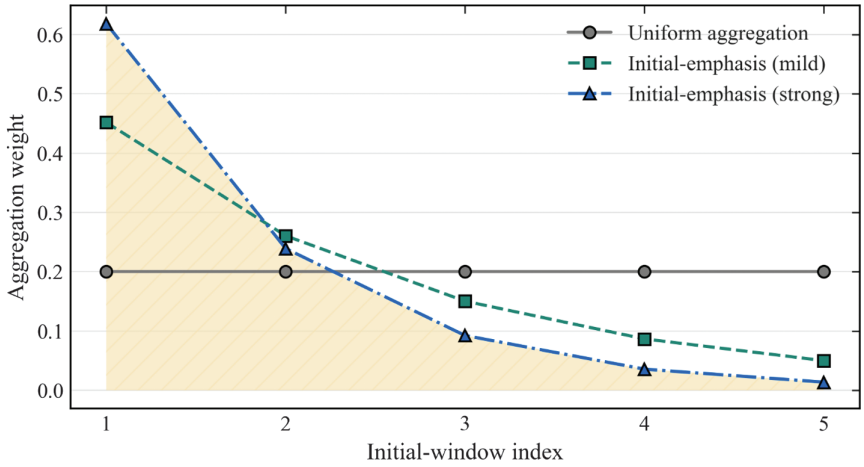

Fig. 3. Prefix-weight profiles for uniform and initial-emphasis aggregation.

III. BAYESIAN PARTIAL POOLING FOR PREFIX-BASED LIFETIME PREDICTION

Once local pseudo-lifetimes are obtained, the central statistical question is how much information should be shared across units. Selective emphasis treats each unit's early windows largely in isolation, whereas partial pooling stabilizes these summaries jointly. The Bayesian formulation addresses three features of the prefix setting: volatility across adjacent windows, heterogeneity across devices, and sensitivity to extreme local extrapolations.

A. HIERARCHICAL MODEL

Define $y_{i,k} = \log L_{i,k}$, the logarithm of the local pseudo-lifetime from window k of unit i. The working observation model is

$$y_{i,k} \sim N(\theta_i, \sigma_w^2) \tag{8}$$

where θ_i is the latent log-lifetime for unit i. Across units, these latent quantities follow

$$\theta_i \sim N(\mu_\theta, \tau^2) \tag{9}$$

μ_θ is the fleet-level mean on the log-lifetime scale, σ_w^2 is the within-device variance among local windows, and τ^2 is the between-device variance. The device-level average of the local log pseudo-lifetimes is denoted by $\bar{y}_i$.

The hierarchy regularizes window instability without forcing all units to share a common lifetime. The within-device variance σ_w^2 measures disagreement among adjacent prefix windows, whereas τ^2 preserves persistent differences among units. The empirical-Bayes implementation estimates $\mu_\theta, \sigma_w^2, \tau^2$ by method of moments on the log pseudo-lifetimes. The fleet mean is the unweighted average of device-level mean log pseudo-lifetimes; the within-device variance is the average local-window variance; and the between-device variance is estimated after subtracting the expected within-device contribution, with a small positive lower bound to avoid degeneracy.

B. SHRINKAGE MECHANISM AND COMPARISON ROLE

The model acts through shrinkage. Units with highly dispersed local-window estimates receive stronger regularization toward the fleet-level center, whereas units with coherent local estimates retain more unit-specific information. The posterior mean has the precision-weighted form:

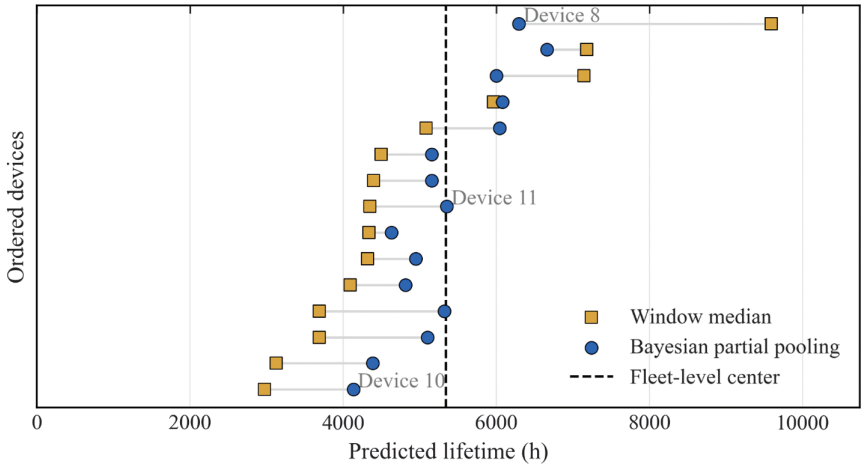

Fig. 4. Window-median and Bayesian partial-pooling lifetime estimates ordered by device.

$$\theta_i^{post} = \alpha_i \bar{y}_i + (1-\alpha_i)\mu_\theta,\ \alpha_i = \frac{K_i\tau^2}{K_i\tau^2 + \sigma_w^2} \qquad (10)$$

where α_i is the data weight implied by the relative within- and between-device precision. The corresponding lifetime predictor is $T_i = \exp(\theta_i^{post})$.

The shrinkage pattern in Fig. 4 shows that pooling moderates local prefix information rather than replacing it. Units with unusually small or large window summaries move toward the fleet-level center while retaining their ordering. Because all estimators use the same prefix information, the comparison isolates the value of cross-unit stabilization relative to selective emphasis of unstable local windows.

IV. PRIMARY CONTROLLED VALIDATION ON THE GaAs LASER DATASET

The empirical work separates two validation roles. The GaAs laser data provide a controlled benchmark in which the same pseudo-lifetime map is applied to complete trajectories and early prefixes, so the effect of prefix treatment is separated from degradation-law selection. Section V then applies the same early-prefix principle to Oxford battery data with observed 95% capacity-retention threshold lives.

The numerical example uses the GaAs laser degradation data reported by Meeker *et al.* [37]. The dataset contains 15 devices, each measured 16 times at 250-hour intervals. The degradation index is expressed as a percentage increase relative to baseline current, and functional failure is defined by the adopted degradation threshold.

A. DATASET, BENCHMARK, AND FIXED-PREFIX PROTOCOL

The early-stage protocol uses only the first six observations of each device for prediction; the remaining observations define a full-history pseudo-lifetime reference. This reference is a model-based projection obtained by applying the same power-law map to all 16 observations of each device. It is not an experimentally observed failure time, because the GaAs laser degradation tests are typically right-censored before the failure threshold is crossed. The evaluation therefore measures agreement with a controlled pseudo-lifetime benchmark rather than physical run-to-failure accuracy.

For consistency between local and global extrapolation, the same pseudo-lifetime map g is used for the full trajectory and for each early-stage window. In the GaAs validation, g is instantiated by the power-law threshold-crossing relations $D(t) = a_i t^{n_i}$ and $T_i = (\frac{D_{fail}}{a_i})^{\frac{1}{n_i}}$. Under this construction, estimator differences are attributed to aggregation and pooling of the same prefix information rather than to model-form selection.

B. MAIN PREFIX-TREATMENT RESULTS

Figure 5 compares device-level lifetime predictions across the competing prefix treatments. Initial-window emphasis produces broad, right-skewed predictions, consistent with the instability of the earliest local extrapolations. Bayesian partial pooling attenuates this instability by shrinking noisy local pseudo-lifetimes toward the fleet-level distribution while retaining device-specific departures. Early global fitting and the window median are more stable than aggressive early weighting, but both remain sensitive to the short observed prefix.

The Supplementary Material gives two diagnostics for the benchmark construction: the adequacy of the power-law map as a common trajectory representation and the distribution of the resulting full-history pseudo-lifetime references on the raw and log scales. These checks document the reference construction without turning the main experiment into a broad model-selection exercise.

The error distributions and bootstrap mean absolute error (MAE) ranking in Figs. 6 and 7 give the same ordering. Bayesian partial pooling has the smallest errors, whereas the initial-emphasis rules have larger and more dispersed errors. The ranking indicates that the gain comes from regularizing unstable local extrapolations rather than from using additional prefix information.

In the GaAs benchmark, Bayesian partial pooling is compared with four pseudo-lifetime comparators and four direct early-feature baselines (Table I). It gives the smallest MAE, 530.9 h. The closest direct-feature competitor is Huber regression, with an MAE of 550.0 h, followed by support vector regression (SVR) at 605.5 h, ridge regression

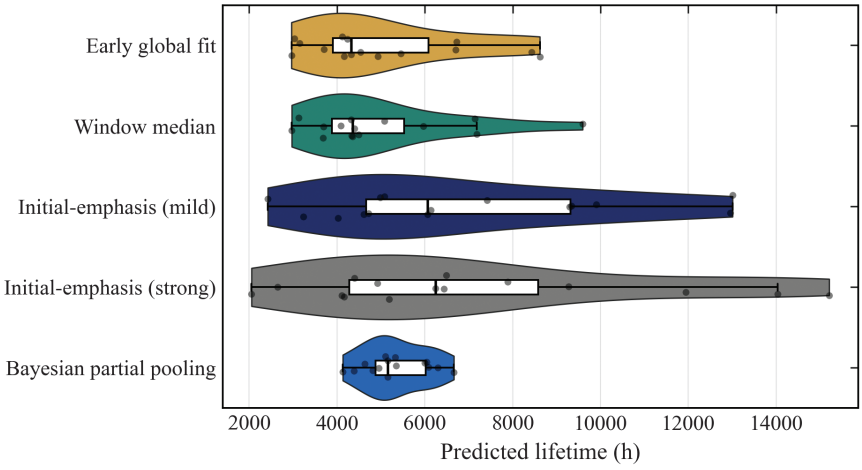

Fig. 5. Device-level lifetime predictions from the five prefix-treatment estimators.

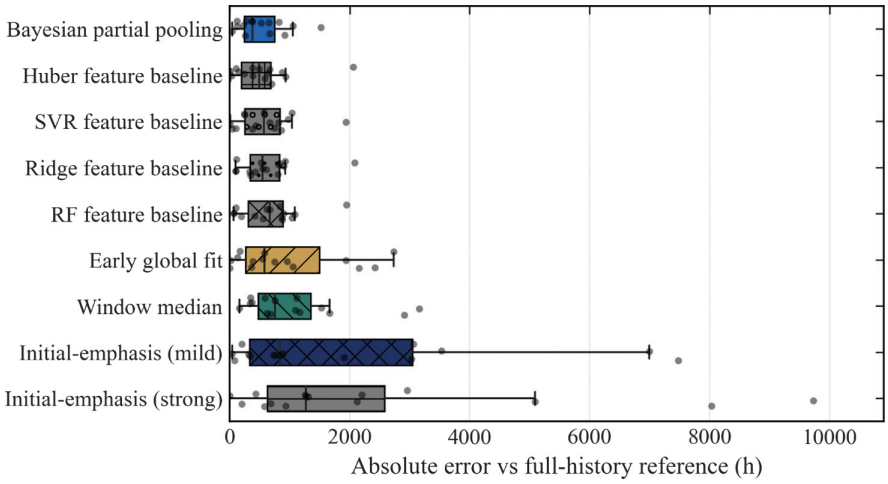

Fig. 6. Absolute prediction errors relative to the full-history pseudo-lifetime reference.

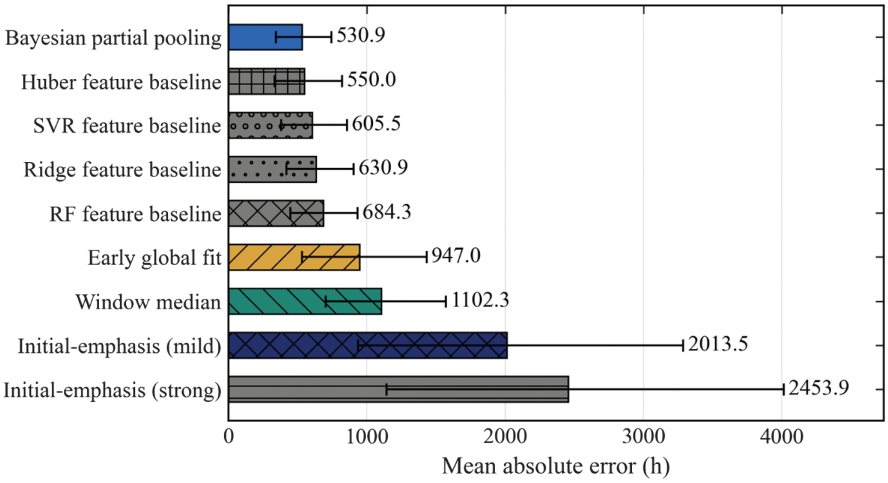

Fig. 7. Mean absolute error with bootstrap 95% confidence intervals under the fixed-prefix protocol.

Table I. GaAs prediction-error comparison relative to the full-history pseudo-lifetime reference; errors are reported in hours

Method	Family	MAE	RMSE	Bias	Median AE	Over-pred. rate
Bayesian partial pooling	Proposed	530.9	663.6	+ 164.2	382.8	60.0%
Huber feature baseline	Direct-feature	550.0	739.3	−122.2	494.0	53.3%
SVR feature baseline	Direct-feature	605.5	774.7	−82.5	573.1	46.7%
Ridge feature baseline	Direct-feature	630.9	792.0	−16.1	545.5	60.0%
RF feature baseline	Direct-feature	684.3	835.2	+ 81.0	668.0	53.3%
Early global fit	Pseudo-lifetime	947.0	1298.5	−168.5	581.6	33.3%
Window median	Pseudo-lifetime	1102.3	1404.7	−212.8	758.4	33.3%
Initial-emphasis (mild)	Pseudo-lifetime	2013.5	3080.7	+ 1704.9	841.4	60.0%
Initial-emphasis (strong)	Pseudo-lifetime	2453.9	3742.3	+ 1826.0	1270.8	60.0%

at 630.9 h, and random forest at 684.3 h. Relative to early global fitting, the strongest pseudo-lifetime comparator at 947.0 h, Bayesian partial pooling reduces MAE by 44.0%; relative to the strongest direct-feature baseline, the reduction is 3.5%. The bias and over-prediction columns retain error direction without imposing an application-specific cost ratio.

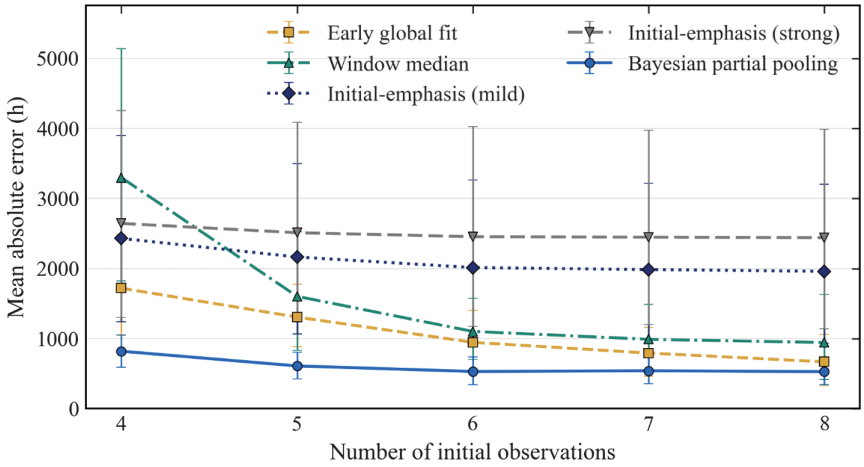

Fig. 8. Mean absolute error across tested prefix lengths.

C. PREFIX-LENGTH SENSITIVITY

Figure 8 tests sensitivity to the amount of available prefix information among the prefix-treatment estimators. Bayesian partial pooling gives the lowest mean absolute error across the plotted prefix lengths, whereas the two initial-emphasis schemes remain the least stable. This pattern indicates that the improvement over local pseudo-lifetime aggregation is not tied to a single arbitrary prefix length.

D. REPRESENTATIVE DEVICE-LEVEL BEHAVIOR

Aggregate error metrics can hide device-level behavior. Figure 9 therefore shows three representative devices spanning shorter-, intermediate-, and longer-lifetime regions of the sample. Across these cases, initial-emphasis weighting tends to shift crossing-time predictions too far to the right, whereas partial pooling moderates local extrapolation without collapsing all devices to a single common prediction.

E. BASELINE AND EMPIRICAL-BAYES SENSITIVITY

The five prefix-treatment strategies introduced in Section 2.D differ only in how the prefix is aggregated; they share the same power-law degradation map and single-stage estimation logic. A frequentist two-stage random-effects power-law model is added as a probabilistic comparator in the spirit of random-coefficient and heterogeneous degradation models [16,17,29,30]. For each device, the prefix is first fitted with a power-law curve, and the log-lifetime implied by that prefix fit is then shrunk through a random-effects Gaussian model before conversion to a pseudo-lifetime prediction.

Figure 10 compares the random-effects power-law model with the five prefix-treatment estimators. Bayesian partial pooling obtains the lowest MAE, 530.9 h; early global fitting and the random-effects power-law baseline yield 947.0 h and 993.3 h, respectively. The two-stage random-effects model improves on the window median and the initial-emphasis rules, but it does not match partial pooling of local pseudo-lifetimes. This result suggests that shrinking a single prefix-fitted lifetime is less effective than regularizing the multiple local extrapolations generated within the prefix. The direct early-feature baselines in Table I provide the complementary comparison outside the pseudo-lifetime framework.

The Bayesian partial-pooling estimator uses three empirical-Bayes hyperparameters: the fleet-level mean log pseudo-lifetime, the within-device variance, and the between-device variance. Because the GaAs sample

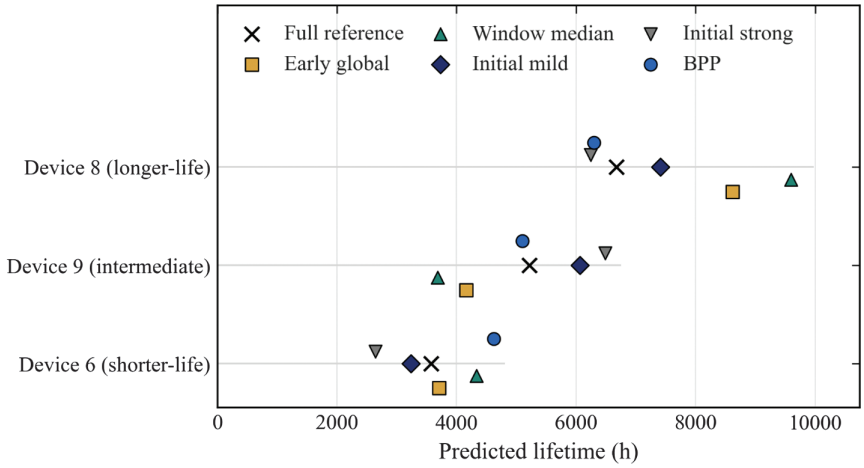

Fig. 9. Device-level predictions for shorter-, intermediate-, and longer-life examples.

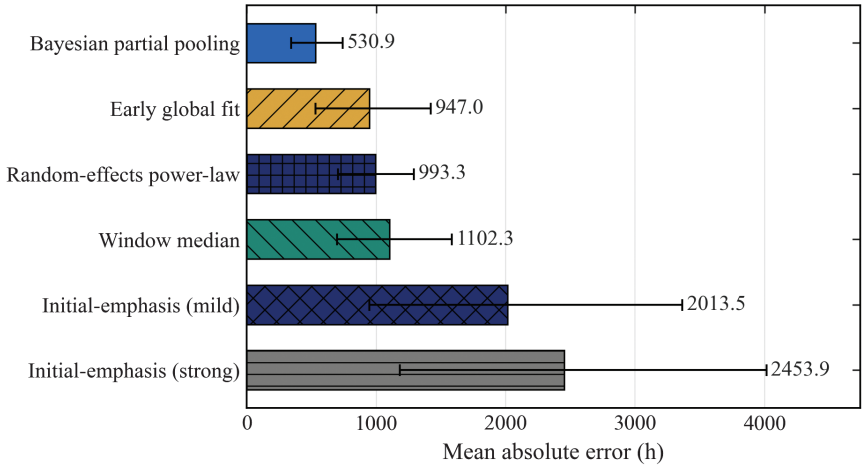

Fig. 10. Mean absolute error for prefix-treatment estimators and the random-effects power-law baseline.

Table II. Empirical-Bayes leave-one-out sensitivity for Bayesian partial pooling

Quantity	Full sample	Leave-one-out (LOO) minimum	LOO maximum
Fleet-level log-lifetime mean	8.574	8.532	8.619
Within-device variance σ^2	0.373	0.349	0.385
Between-device variance τ^2	0.048	0.023	0.063
Population shrinkage weight	0.606	0.527	0.770

contains only 15 devices, these estimates may be sensitive to extreme devices. Table II reports a leave-one-out sensitivity analysis. For the full sample, the estimated population shrinkage weight is 0.606. Across the 15 leave-one-out refits, it ranges from 0.527 to 0.770, with a mean of 0.611 and a standard deviation of 0.064. The fleet-level mean log pseudo-lifetime and the variance components show similarly bounded variation. These ranges indicate that the partial-pooling behavior is not determined by a single device.

The pooled estimator is also stable across leave-one-out folds: devices with high shrinkage under the full-sample model retain high shrinkage in every refit, and the rank ordering of pooled lifetime predictions is unchanged. This stability limits the influence of any single device in the empirical-Bayes step.

F. UNCERTAINTY AND RESIDUAL DIAGNOSTICS

Point error metrics summarize accuracy but do not assess calibration or residual structure. Posterior credible intervals and residual diagnostics are therefore used to evaluate Bayesian partial pooling beyond point prediction.

The per-device 90% and 95% posterior credible intervals in Fig. 11 have empirical coverage of 100% at both nominal levels, with median widths of 2945 h and 3529 h, respectively. The intervals are slightly conservative in this small sample, consistent with an empirical-Bayes

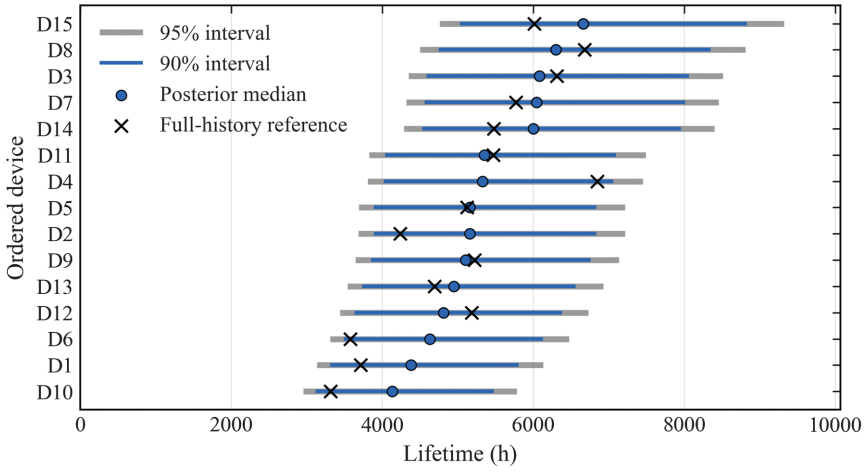

Fig. 11. Bayesian partial-pooling posterior intervals and full-history pseudo-lifetime references.

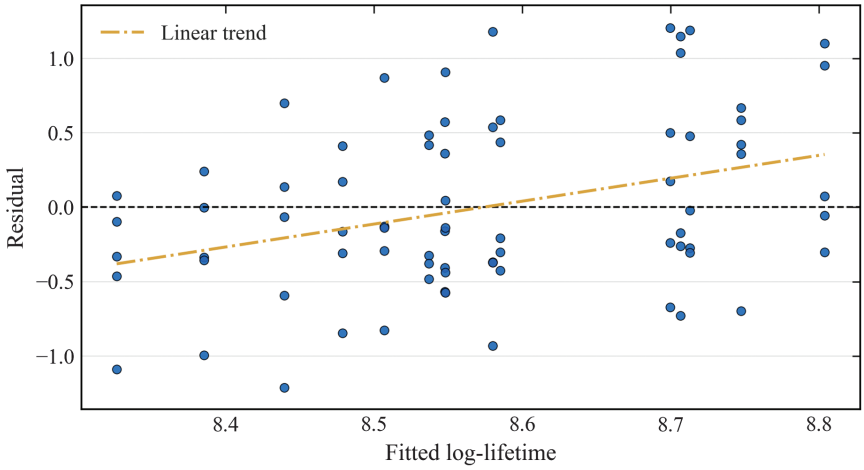

Fig. 12. Log-scale residuals plotted against posterior fitted log-lifetime.

implementation that does not propagate hyperparameter uncertainty. They nevertheless provide an uncertainty statement for each early-stage prediction rather than only a point estimate.

The Bayesian partial-pooling model assumes that, conditional on the latent device-level log-lifetime, local log pseudo-lifetimes are Gaussian with constant within-device variance. This homoscedasticity assumption yields a closed-form shrinkage update but requires diagnostic assessment in a 15-device sample. Figure 12 plots log-scale residuals against fitted log-lifetimes, with a fitted linear trend used to check whether residual location changes systematically across the prediction range.

The residual mean is −3.79e-16 on the log scale, and the residual standard deviation is 0.587. Bartlett's test of equal variances across local windows gives $p = 0.112$, and Levene's robust test gives $p = 0.095$. Neither test rejects constant variance at the 5% level, although the Levene result is borderline. The residual scatter remains centered near zero without a strong fan-shaped pattern, supporting the homoscedastic working model while motivating heteroscedastic extensions for larger datasets.

V. EXTERNAL VALIDATION ON OXFORD LITHIUM-ION BATTERY CELLS

A. PUBLIC BATTERY DATASET AND FIXED EARLY-CYCLE PROTOCOL

The Oxford battery study complements the GaAs benchmark by retaining the fixed early-information protocol and baseline grouping while evaluating predictions against observed threshold lives. The public Oxford Battery Degradation Dataset [38,39] contains eight Kokam pouch cells cycled under a common protocol, with periodic 1C discharge-capacity characterization. The validation target is the first interpolated cycle at which capacity falls below 95% of the cell-specific initial reference capacity. This early health-threshold target is used because all eight cells cross it within the public record, whereas a conventional 80% capacity threshold is not observed for every cell. The observed threshold lives range from 1040.6 to 1706.5 cycles.

The eight discharge-capacity trajectories in Fig. 13 share a common degradation direction but differ in

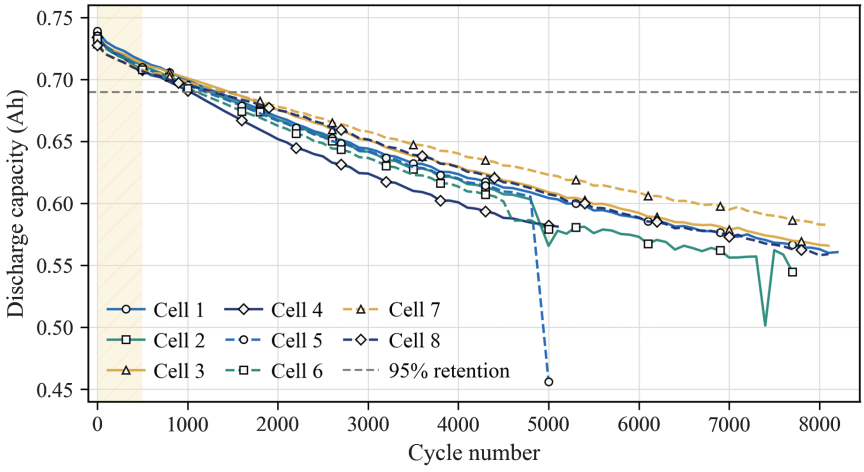

Fig. 13. Oxford battery discharge-capacity trajectories with the first-500-cycle prediction window and 95% capacity-retention threshold.

threshold-crossing time. This structure makes the dataset suitable for testing whether prefix-derived local lifetimes can be stabilized across comparable cells under a fixed first-500-cycle information budget.

B. EXTERNAL THRESHOLD-LIFE PREDICTION UNDER A FIXED EARLY-CYCLE BUDGET

Predictions use only the first 500 cycles. Adjacent 100-cycle windows are converted into local pseudo-lifetimes and calibrated within each leave-one-out training fold. Bayesian partial pooling combines the calibrated local lifetime estimates using window-specific reliability, so noisier windows contribute less to the posterior cell-level lifetime. The comparison retains four pseudo-lifetime alternatives: window median, early global mean, mild initial emphasis, and strong initial emphasis. It also includes four direct first-prefix regressors: Huber regression, ridge regression, random forest, and radial-kernel SVR. These regressors use first-500-cycle capacity summaries, including selected capacities, capacity drops, slope, curvature, and dispersion, and provide stricter comparators outside the proposed pseudo-lifetime aggregation framework.

Under this observed threshold-life endpoint, Bayesian partial pooling achieves the lowest leave-one-out MAE, 21.0 cycles (Table III and Fig. 14). The window-median pseudo-lifetime comparator follows at 38.9 cycles, and the strongest direct-feature baseline is ridge regression at 79.2 cycles. Relative to these two strongest alternatives, Bayesian partial pooling reduces MAE by 46.0% and 73.5%, respectively. The bias and over-prediction-rate columns quantify error direction, and the battery validation extends the comparison from a model-derived GaAs reference to an observed threshold-life task.

C. COMPONENT ABLATION AND PREDICTIVE UNCERTAINTY

The component ablations in Table IV clarify which parts of the battery model carry the accuracy gain. Removing fleet-level shrinkage increases MAE only from 21.0 to 21.4 cycles. In contrast, removing partial pooling, window-specific reliability, or local-window calibration increases MAE to 38.9, 63.4, and 68.0 cycles, respectively. The battery improvement is therefore driven mainly by calibrated local lifetimes and reliability-weighted pooling, with fleet-level shrinkage providing additional stabilization in the eight-cell sample.

Cell-level predictions in Fig. 15 remain close to the ideal-agreement line for most batteries. The largest absolute error is 60.4 cycles, and the median absolute error is 10.7 cycles. This

Table III. Oxford battery validation errors for observed 95% capacity-retention threshold lives under leave-one-out validation

Method	Family	MAE	RMSE	Bias	Median absolute error (AE)	Over-pred. rate
Bayesian partial pooling	Proposed	21.0	28.8	+5.6	10.7	75.0%
Window median	Pseudo-lifetime	38.9	50.9	+0.9	30.9	75.0%
Early global mean	Pseudo-lifetime	61.7	74.0	+3.9	58.6	62.5%
Ridge feature baseline	Direct-feature	79.2	118.4	−10.2	49.7	37.5%
Huber feature baseline	Direct-feature	90.9	114.6	−4.4	69.7	50.0%
Initial-emphasis mild	Pseudo-lifetime	95.9	122.3	+14.2	84.9	62.5%
Random forest (RF) feature baseline	Direct-feature	105.1	155.8	−31.4	41.4	50.0%
Initial-emphasis strong	Pseudo-lifetime	119.4	160.9	+20.5	98.7	62.5%
SVR feature baseline	Direct-feature	167.0	201.5	−9.4	129.3	62.5%

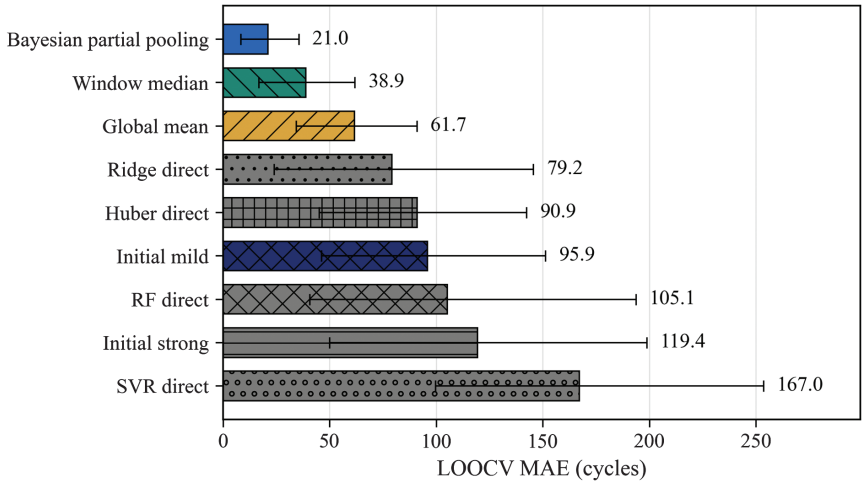

Fig. 14. Oxford battery leave-one-out cross-validation (LOOCV) mean absolute error with bootstrap 95% confidence intervals and hatching for estimator families.

Table IV. Oxford battery sensitivity and ablation for Bayesian partial pooling under leave-one-out validation

Configuration	Perturbation	MAE	Delta MAE	Bias	Over-pred. rate
Full Bayesian partial pooling	Complete model	21.0	+0.0	+5.6	75.0%
No fleet shrinkage	Remove population prior	21.4	+0.4	+6.1	75.0%
No partial pooling	Window median	38.9	+17.9	+0.9	75.0%
Equal window reliability	Use common window variance	63.4	+42.4	−2.4	62.5%
No local-window calibration	Use raw local extrapolations	68.0	+46.9	−10.9	62.5%

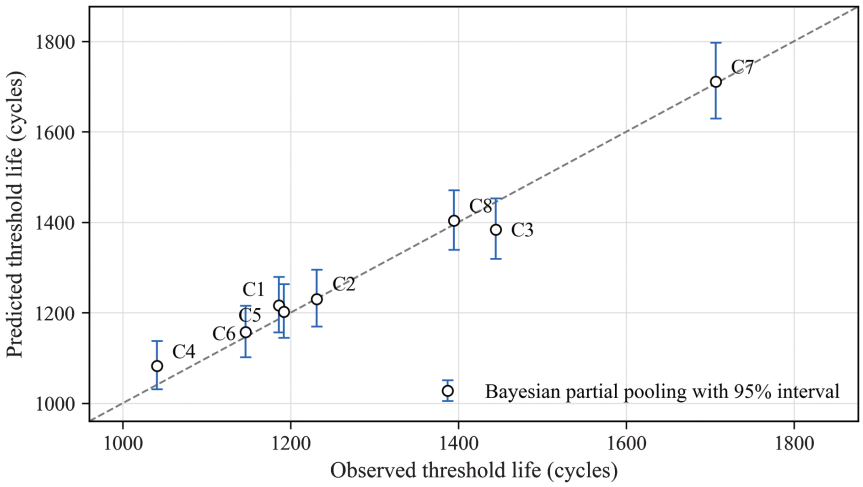

Fig. 15. Predicted and observed Oxford battery threshold lives with 95% predictive intervals.

agreement is consistent with the proposed mechanism: early local extrapolations contain useful ordering information, but their window-to-window volatility must be attenuated before a stable lifetime forecast can be obtained.

The predictive intervals in Fig. 15 provide a compact uncertainty assessment. The nominal 90% and 95% intervals cover 7 of 8 and 8 of 8 held-out cells, corresponding to empirical coverages of 87.5% and 100.0%, respectively. The median interval widths are 104.1 and 124.0 cycles. Although the validation sample is small, these coverage results indicate that the low MAE is not obtained through overconfident prediction.

VI. CONCLUSION

Prefix treatment is a central design choice in early-stage lifetime prediction from degradation data. The GaAs laser benchmark isolates this effect under a fixed pseudo-lifetime map, and the Oxford battery validation evaluates the same principle against observed threshold lives. In both settings, short local extrapolations are informative, but they require regularization before they can support reliable lifetime prediction.

Bayesian partial pooling provides this regularization by shrinking noisy local pseudo-lifetimes while retaining device-level departures. In the GaAs benchmark, it gives the lowest MAE among the evaluated methods, including pseudo-lifetime comparators and direct-feature baselines, and its posterior intervals cover the full-history pseudo-lifetime references. In the Oxford battery validation, it gives the lowest leave-one-out MAE among the evaluated methods, with a 46.0% reduction relative to the strongest pseudo-lifetime comparator and a 73.5% reduction relative to the strongest direct-feature baseline. The Oxford ablation further indicates that partial pooling, window-specific reliability, and local-window calibration are the main contributors to the accuracy gain, while the nominal 95% predictive intervals retain 100.0% empirical coverage.

The conclusions are bounded by the intended scope of the method. The GaAs reference is a full-history pseudo-lifetime projection rather than a directly observed physical failure time, and both empirical datasets are small: 15 GaAs devices and eight Oxford battery cells. The empirical-Bayes implementation also uses a log-normal working likelihood and plug-in hyperparameter estimates. Further work should examine fully Bayesian hyperparameter propagation, heteroscedastic window-level variance, broader battery and laser datasets, and physics-informed degradation constraints. Within this scope, the comparative evidence is consistent for the primary accuracy endpoints: when comparable degradation units provide noisy but informative early local extrapolations, Bayesian partial pooling offers a more accurate, interpretable, and uncertainty-aware prefix treatment than heuristic aggregation or direct first-prefix regression.

ACKNOWLEDGMENTS

This research was supported by the Jilin Provincial Scientific and Technological Development Program under Grant No. 20260205059GH.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

SUPPLEMENTARY MATERIAL

PURPOSE OF THIS SUPPLEMENT

Two checks support the main article. The first examines whether the common power-law map used to define the GaAs pseudo-lifetime reference is a reasonable monotone representation for the laser degradation trajectories. The second characterizes the distributional shape of the full-history pseudo-lifetime references derived from that map. Both checks are descriptive and support the modeling choices used in the main article; they are not treated as the source of the methodological contribution.

SA. SUPPORT FOR THE COMMON POWER-LAW MAP

The main analysis fixes a common power-law degradation map to isolate prefix-treatment effects rather than to conduct broad model selection. This choice still requires an adequacy check. Five monotone trend models are fitted to each complete degradation trajectory: linear, quadratic, power-law, exponential, and logarithmic. The comparison addresses a narrow question: whether the power-law map is a reasonable common monotone approximation for defining a full-history pseudo-lifetime reference in this dataset.

In the device-level comparison in Table SI, the power-law model gives the smallest mean Akaike information criterion (AIC) among the five candidate trend laws and the smallest median root mean square error (RMSE). It is tied with the logarithmic model for the largest number of device-level best-AIC selections. This result does not imply that the power-law model is uniquely correct; it indicates that the adopted map is sufficiently competitive to serve as a common pseudo-lifetime definition without making model-form selection the dominant issue. Figure S1 provides the corresponding fleet-level visual check for the fitted monotone trend laws.

Table SI. Summary of monotone trend-model comparison over the 15 complete degradation trajectories

Model	Mean AIC	Mean Bayesian information criterion (BIC)	Median RMSE	Best-AIC count
Power-law	−59.16	−57.61	0.133	5
Logarithmic	−55.19	−53.65	0.168	5
Quadratic	−52.66	−51.12	0.165	1
Linear	−52.25	−51.47	0.191	4
Exponential	−28.19	−26.64	0.399	0

SB. DISTRIBUTIONAL DIAGNOSTICS FOR THE REFERENCE LIFETIME SCALE

After the common power-law map is fixed, each complete GaAs degradation trajectory yields a full-history pseudo-lifetime reference. The main analysis does not require selection of a single parametric lifetime distribution for this sample; the reference lifetimes serve only as a common benchmark for early-stage prediction errors. Their distributional shape is documented because Bayesian partial pooling is implemented on the log-lifetime scale.

Table SII reports an exploratory distributional screen over Weibull, Lognormal, Gamma, Nakagami, Log-logistic, Rice, and Rayleigh distributions. The Rayleigh fit gives the smallest AIC in this small sample because of its parsimony, whereas the Weibull fit gives the best Kolmogorov–Smirnov alignment. Differences among the leading candidates remain modest. The diagnostic role of this

Table SII. Distributional diagnostics for the full-history pseudo-lifetime reference sample

Distribution	AIC	BIC	Kolmogorov-Smirnov (KS) statistic	KS p-value
Rayleigh	255.85	257.27	0.204	0.497
Weibull	257.04	259.17	0.116	0.974
Rice	257.57	259.69	0.147	0.859
Nakagami	257.76	259.89	0.162	0.772
Gamma	258.17	260.30	0.175	0.685
Lognormal	258.71	260.83	0.189	0.594
Log-logistic	259.47	261.60	0.153	0.825

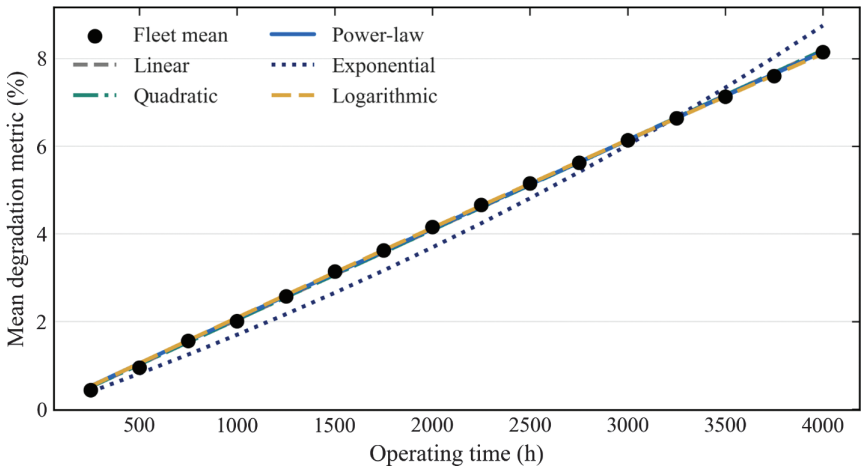

Fig. S1. Fleet-mean degradation trajectory with fitted monotone trend models.

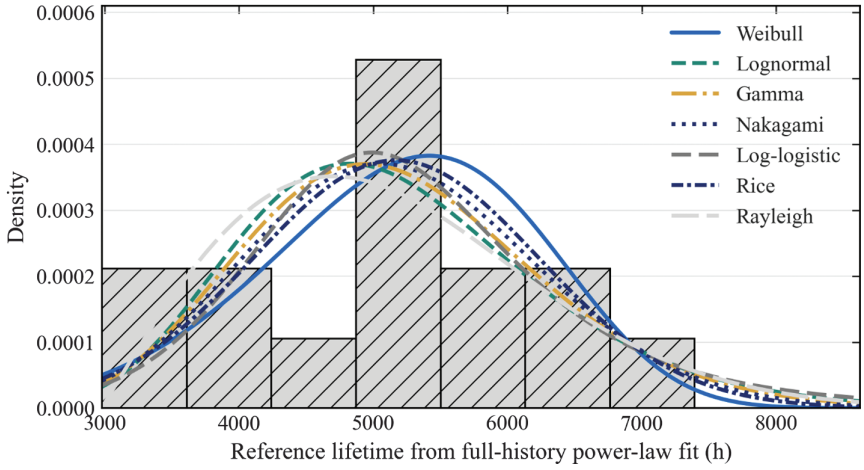

Fig. S2. Candidate distribution fits for the full-history pseudo-lifetime references.

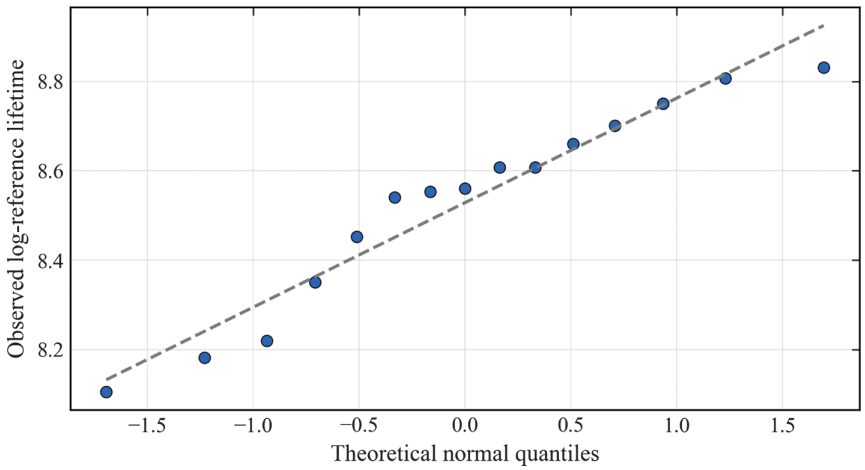

Fig. S3. Normal Q-Q plot for log-transformed reference lifetimes.

analysis is descriptive: the reference lifetimes are positive and right-skewed, and the log transformation used in the partial-pooling model is therefore reasonable. The fitted density overlays in Figure S2 and the normal Q-Q plot in Figure S3 provide complementary visual checks.

REFERENCES

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] X.-S. Si *et al.*, "Remaining useful life estimation: A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.

[3] Y. Lei *et al.*, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018.

[4] JDMD Editorial Office, N. Gebraeel *et al.*, "Prognostics and remaining useful life prediction of machinery: Advances, opportunities and challenges," *J. Dyn. Monit. Diagn.*, vol. 2, no. 1, pp. 1–12, 2023.

[5] P. Luo *et al.*, "Research on remaining useful life prediction method of rolling bearing based on health indicator extraction and trajectory enhanced particle filter," *J. Dyn. Monit. Diagn.*, vol. 1, no. 2, pp. 66–83, 2022.

[6] A. Nagheli, M. Poursina, and H. Karimpour, "A Bayesian prognosis framework for rolling bearings based on total harmonic distortion health indicator and nonlinear Wiener process," *J. Dyn. Monit. Diagn.*, vol. 5, no. 1, pp. 74–85, 2026.

[7] M. Singh, "Power prediction from wind turbine SCADA data in the presence of modeling and measurement uncertainties," *J. Dyn. Monit. Diagn.*, vol. 4, no. 4, pp. 274–291, 2025.

[8] H. Li *et al.*, "A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 209, p. 111120, 2024.

[9] J. Zhang *et al.*, "An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty," *Reliab. Eng. Syst. Saf.*, vol. 222, p. 108357, 2022.

[10] X. Shao *et al.*, "Data-model-linked remaining useful life prediction method with small sample data: A case of subsea valve," *Reliab. Eng. Syst. Saf.*, vol. 250, p. 110323, 2024.

[11] J. Pan *et al.*, "Probabilistic remaining useful life prediction without lifetime labels: A Bayesian deep learning and stochastic process fusion method," *Reliab. Eng. Syst. Saf.*, vol. 250, p. 110313, 2024.

[12] C. Lopez-Salazar *et al.*, "A comprehensive framework for estimating the remaining useful life of Li-ion batteries under limited data conditions with no temporal identifier," *Reliab. Eng. Syst. Saf.*, vol. 253, p. 110517, 2025.

[13] Y. Dong *et al.*, "An interpretable multiscale lifting wavelet contrast network for planetary gearbox fault diagnosis with small samples," *Reliab. Eng. Syst. Saf.*, vol. 251, p. 110404, 2024.

[14] D. Gao *et al.*, "Semi-supervised small sample fault diagnosis under a wide range of speed variation conditions based on uncertainty analysis," *Reliab. Eng. Syst. Saf.*, vol. 242, p. 109746, 2024.

[15] R. Liu *et al.*, "Knowledge-informed FIR-based cross-category filtering framework for interpretable machinery fault diagnosis under small samples," *Reliab. Eng. Syst. Saf.*, vol. 254, p. 110610, 2025.

[16] W. Peng *et al.*, "Inverse Gaussian process models for degradation analysis: A Bayesian perspective," *Reliab. Eng. Syst. Saf.*, vol. 130, pp. 175–189, 2014.

[17] T. Yuan and Y. Ji, "A hierarchical bayesian degradation model for heterogeneous data," *IEEE Trans. Reliab.*, vol. 64, no. 1, pp. 63–70, 2015.

[18] X. Dai *et al.*, "Reliability modelling of wheel wear deterioration using conditional bivariate gamma processes and Bayesian hierarchical models," *Reliab. Eng. Syst. Saf.*, vol. 226, p. 108710, 2022.

[19] R. Chen *et al.*, "Importance measures for critical components in complex system based on Copula hierarchical Bayesian network," *Reliab. Eng. Syst. Saf.*, vol. 230, p. 108883, 2023.

[20] T. Huang *et al.*, "Merging multi-level evidential observations for dynamic reliability assessment of hierarchical multi-state systems: A dynamic Bayesian network approach," *Reliab. Eng. Syst. Saf.*, vol. 249, p. 110225, 2024.

[21] Y. Yu, B. Shuai, and W. Huang, "Resilience evaluation of train control on-board system based on multi-dimensional continuous-time Bayesian network," *Reliab. Eng. Syst. Saf.*, vol. 246, p. 110099, 2024.

[22] P. G. Morato *et al.*, "Inference and dynamic decision-making for deteriorating systems with probabilistic dependencies through Bayesian networks and deep reinforcement learning," *Reliab. Eng. Syst. Saf.*, vol. 235, p. 109144, 2023.

[23] L. A. Bull *et al.*, "Hierarchical Bayesian modeling for knowledge transfer across engineering fleets via multitask learning," *Comput.-Aided Civ. Infrastruct. Eng.*, vol. 38, no. 7, pp. 821–848, 2023.

[24] A. Xu *et al.*, "Bayesian reliability assessment of permanent magnet brake under small sample size," *IEEE Trans. Reliab.*, vol. 74, no. 1, pp. 2107–2117, 2025.

[25] H. Shen *et al.*, "Hierarchical Bayesian support vector regression with model parameter calibration for reliability modeling and prediction," *Reliab. Eng. Syst. Saf.*, vol. 229, p. 108842, 2023.

[26] M. Mishra *et al.*, "Bayesian hierarchical model-based prognostics for lithium-ion batteries," *Reliab. Eng. Syst. Saf.*, vol. 172, pp. 25–35, 2018.

[27] E. Vanem *et al.*, "Statistical models for condition monitoring and state of health estimation of lithium-ion batteries for ships," *J. Dyn. Monit. Diagn.*, vol. 3, no. 1, pp. 11–20, 2024.

[28] S. K. Thakuri *et al.*, "The RUL prediction of Li-ion batteries based on adaptive LSTM," *J. Dyn. Monit. Diagn.*, vol. 4, no. 1, pp. 53–64, 2025.

[29] Z. Wang, Q. Zhai, and P. Chen, "Degradation modeling considering unit-to-unit heterogeneity-A general model and comparative study," *Reliab. Eng. Syst. Saf.*, vol. 216, p. 107897, 2021.

[30] H. Hao, Z. Ji, and C. Li, "Heterogeneous Degradation Modeling Based on Hierarchical Bayesian Model and Wiener Process," *Iran. J. Sci.*, vol. 47, no. 2, pp. 457–466, 2023.

[31] G. A. Veloso and R. H. Loschi, "Dynamic linear degradation model: Dealing with heterogeneity in degradation paths," *Reliab. Eng. Syst. Saf.*, vol. 210, p. 107446, 2021.

[32] C. C. Santos and R. H. Loschi, "Semi-parametric Bayesian models for heterogeneous degradation data: An application to laser data," *Reliab. Eng. Syst. Saf.*, vol. 202, p. 107038, 2020.

[33] D. Liu *et al.*, "Bayesian model averaging based reliability analysis method for monotonic degradation dataset based on inverse Gaussian process and Gamma process," *Reliab. Eng. Syst. Saf.*, vol. 180, pp. 25–38, 2018.

[34] F. Duan and G. Wang, "Bayesian analysis for the transformed exponential dispersion process with random effects," *Reliab. Eng. Syst. Saf.*, vol. 217, p. 108104, 2022.

[35] X. Cao and K. Peng, "Stochastic uncertain degradation modeling and remaining useful life prediction considering aleatory and epistemic uncertainty," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[36] F. Safaei and S. Taghipour, "Integrated degradation-based burn-in and maintenance model for heterogeneous and highly reliable items," *Reliab. Eng. Syst. Saf.*, vol. 244, p. 109942, 2024.

[37] W. Q. Meeker, L. A. Escobar, and F. G. Pascual, *Statistical Methods for Reliability Data*, 2nd ed. Hoboken, NJ, USA: John Wiley & Sons, 2021.

[38] D. Howey and C. Birkl, The Oxford Battery Degradation Dataset 1, University of Oxford, 2017.

[39] C. Birkl, "Diagnosis and prognosis of degradation in lithium-ion batteries," DPhil thesis, University of Oxford, 2017.