



Pseudo-label augmentation with rated-voltage covariate for Bayesian lifetime estimation of high-voltage insulation from scarce and censored ALT data

Sewon Kim ^a, Donghwan Kim ^a, Joongki Oh ^b, Hyunsoo Yoon ^a,*

^a Department of Industrial Engineering, Yonsei University, Seoul, 03722, Republic of Korea

^b HD Hyundai Electric, Reliability Assessment Center, Yongin-si, Gyeonggi-do, 16891, Republic of Korea

ARTICLE INFO

Keywords:

Insulation system
Reliability
Accelerated life test
Bayesian estimation
Pseudo-labeling

ABSTRACT

High-voltage rotating-machines frequently fail due to insulation degradation; however, accelerated life-test (ALT) data are scarce and heavily censored. Moreover, conventional ALT models analyze each voltage class independently, which limits cross-class extrapolation. To address these challenges, this study develops **RIPPLE** (Rated-voltage-Informed Pseudo-Probabilistic Lifetime Estimation), a Bayesian accelerated failure-time framework that integrates experimental data and specification information to improve lifetime inference. First, censored specimens are transformed into pseudo-labeled lifetimes by learning a mapping from diagnostic time series and design features to failure time using a lightweight Transformer-MLP model ensemble, with predictive uncertainty explicitly quantified. Second, rated-voltage R is incorporated as a design covariate in a LogNormal rated-voltage-informed Temperature-Nonthermal (R -TNT) stress-life model, enabling joint modeling across heterogeneous voltage classes. Third, pseudo-labels are treated as uncertain soft evidence through a truncated latent-variable likelihood, propagating prediction uncertainty into the posterior and enforcing censoring consistency probabilistically. Applied to multi-stress ALT data on 6.6/13.8 kV form-wound coils, RIPPLE reduced the 90% credible-interval width by approximately 90% at 155 °C and 2.1 R relative to the censored-likelihood baseline, while substantially improving point-prediction accuracy (MAPE 0.453 vs. 2.213). Simulation studies further support the effectiveness of the proposed framework.

1. Introduction

High-voltage rotating-machines, including motors and generators, are essential assets in power generation, manufacturing, and marine operations where stable performance and long-term reliability are critical. These machines operate continuously under high voltage and elevated temperature, conditions that render the insulation system the most vulnerable component. Failures of stator insulation are a well-recognized cause of unexpected outages, leading to costly downtime, increased maintenance demands, and production losses [1–3]. Accordingly, the quantitative evaluation and prediction of insulation lifetime are indispensable for ensuring system reliability and enabling effective preventive maintenance. Accelerated life testing (ALT) is therefore widely adopted to address this challenge. In ALT, specimens are subjected to elevated stress conditions so that degradation and failure occur within a shorter time, enabling extrapolation to service conditions [4,5].

Nevertheless, lifetime data from ALT for rotating-machine insulation systems remain extremely scarce. Insulation degradation proceeds

slowly, and even under accelerated conditions, test durations often exceed one year due to practical limitations on stress acceleration. In principle, ALT stress levels should be selected to ensure failure occurs within a target duration (e.g., 5000 h); however, stresses must remain within acceleration regimes that preserve representative aging mechanisms, rather than inducing non-physical material breakdown [6]. For rotating-machine insulation, which comprises multiple materials such as mica, fibers, and resin, the admissible stress range is further constrained, as excessive thermal or electrical stress may lead to premature failure of the resin rather than gradual insulation aging [7]. Consequently, stress levels cannot be arbitrarily increased to shorten test duration, resulting in prolonged experiments and a high proportion of right-censored observations [8,9].

In addition, the number of parallel tests is inherently limited. Although increasing the number of specimens would improve the availability of lifetime data by yielding more failures, practical constraints restrict such scaling. Combined high-voltage and high-temperature testing requires specialized chambers with stringent thermal uniformity

* Corresponding author.

E-mail addresses: sewon3397@yonsei.ac.kr (S. Kim), cain0709@yonsei.ac.kr (D. Kim), oh.joongki@hd.com (J. Oh), hs.yoon@yonsei.ac.kr (H. Yoon).

<https://doi.org/10.1016/j.ress.2026.113005>

Received 7 January 2026; Received in revised form 3 June 2026; Accepted 7 June 2026

Available online 12 June 2026

0951-8320/© 2026 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

and sufficient insulation clearance, while each specimen must be manufactured on an actual production line, incurring substantial cost and limiting sample availability. As a result, only a few failures are typically observed, rendering extrapolation to real operating conditions highly uncertain [10,11].

To address the limitations of failure-based ALT under data scarcity, accelerated degradation testing (ADT) has been proposed, wherein diagnostic indicators are monitored over time, and failure is defined as the moment when a signal exceeds a predefined threshold [12]. However, for insulation systems, commonly used signals such as dielectric loss tangent ($\Delta \tan \delta$) and insulation resistance are often non-monotonic and noisy, and are further affected by measurement noise and environmental variability [13–15]. As a result, their temporal trajectories cannot be consistently mapped to physical deterioration or failure, which limits the effectiveness of conventional ADT modeling approaches [16]. Furthermore, in practical rotating-machine insulation testing, diagnostic measurements are performed offline between aging sub-cycles rather than continuously during operation, and often require the machine to be taken out of service and tested under controlled conditions [1]. These practical limitations further restrict the direct applicability of conventional ADT frameworks.

Beyond these measurement challenges, most existing reliability studies have focused on simplified material specimens or polymeric cables rather than complete stator insulation systems. Empirical data for large rotating-machine insulation are rare, and the relationship between accelerated aging and in-service lifetime remains uncertain [17]. In practice, insulation degradation is driven by the combined effects of thermal and electrical stresses, as load-induced heating and applied voltages interact continuously. Accordingly, international standards prescribe ALT protocols with thermal [18], electrical [19], or combined thermal–electrical stresses [20]. To better reflect such coupled operating conditions, multi-stress accelerated life modeling has been actively studied to capture interaction effects between stresses under realistic operating environments [15,21,22]. In addition, corresponding test design methodologies have been developed to improve estimation efficiency while maintaining physically meaningful degradation mechanisms [23]. However, practical implementation of combined-stress ALT for rotating-machine insulation systems remains limited due to the scarcity of empirical lifetime data under realistic coupled stresses. Consequently, most reliability models continue to be predominantly calibrated using single-stress datasets, such as those based on the Arrhenius law for thermal aging [4], or the inverse power law (IPL) for electrical stress [5,24].

Another important challenge is design diversity. Insulation systems vary in geometry and material composition, resulting in distinct lifetime distributions and acceleration parameters. As a consequence, no single model can adequately capture all product variations, and most ALT studies therefore focus on a single product type [25,26]. Notably, the rated-voltage (R) of a machine represents more than a nominal specification. Insulation structures are designed for each voltage rating to achieve optimal efficiency, and R therefore correlates strongly with inherent insulation strength [17,27,28]. R serves as a class-level proxy for insulation design margin, reflecting key attributes such as material thickness, conductor spacing, and reinforcement level [29].

These challenges—limited failures, heavy censoring, measurement noise, and heterogeneous design conditions—have motivated the adoption of probabilistic and data-driven approaches that explicitly quantify uncertainty. Bayesian inference provides a principled framework for handling censored data and incorporating prior knowledge, and has been widely applied to fuse accelerated testing data with condition monitoring or field data for improved reliability estimation [30,31]. In many practical scenarios, complete lifetime labels are unavailable, and only indirect information from diagnostic or condition monitoring data can be obtained. Recent studies have addressed this challenge by generating surrogate labels or probabilistic lifetime estimates from such data, while explicitly accounting for their inherent uncertainty [32,33].

These surrogate labels, often referred to as pseudo-labels, are inherently noisy and ambiguous, as they are derived from imperfect models or incomplete observations. Consequently, treating pseudo-labels as exact failure observations can lead to overconfident and biased inference, highlighting the need to propagate their uncertainty into the reliability modeling framework.

Within this context, this study proposes **RIPPLE** (Rated-voltage-Informed Pseudo-Probabilistic Lifetime Estimation), a Bayesian accelerated failure-time framework that enables reliable lifetime prediction and cross-class extrapolation under severe data scarcity. RIPPLE integrates three key methodological components: (i) deep learning-based **pseudo-labeling** to infer failure-like labels for censored samples using diagnostic and design features, thereby effectively increasing the informative sample size when direct failure observations are insufficient; (ii) a **rated-voltage-informed T-NT (R-TNT)** lifetime model that extends the standard T-NT formulation by explicitly incorporating R as a class-level design covariate, enabling cross-class generalization across heterogeneous voltage classes; and (iii) two **probabilistic uncertainty-handling** formulations for pseudo-labeled observations—a variance-inflation likelihood and a truncated latent likelihood—both designed to treat pseudo-labels as soft evidence rather than exact targets, with the truncated latent likelihood adopted as the primary approach based on its superior empirical performance. Accordingly, RIPPLE adopts a Bayesian inference framework to incorporate prior knowledge and support stable estimation under limited and noisy data conditions.

2. Related works

2.1. Industrial context and standards

Extensive reliability surveys on large motors and generators have consistently identified stator insulation failure as a dominant cause of breakdown [1,34]. Consequently, the reliability of insulation systems has long been a critical concern in industrial practice. To establish a consistent testing framework, the IEC 60034-18 series specifies ALT procedures under thermal (–31), electrical (–32), and combined thermal–electrical (–33) stresses [18–20]. Although IEC/TS 60034-18-33 provides guidance for endurance testing under simultaneous multi-factor stresses, it does not define a universal life-stress relationship that can accommodate realistic design variations. In practical industrial environments, accelerated tests frequently yield only a limited number of failures and a high proportion of right-censored observations, which fundamentally constrains statistical inference and increases uncertainty in lifetime estimation. In addition, recent industrial reports have highlighted the high cost of fabricating large coils and the limited chamber capacity required for combined high-voltage and high-temperature testing [10,11].

2.2. Multi-stress lifetime models

Classical single-stress formulations include the Arrhenius model for thermal degradation [4] and the IPL for electrical endurance [5,24]. In practical operating environments, however, insulation systems are subjected to concurrent thermal and electrical stresses, which has motivated the development of multi-stress lifetime models. To capture such coupled degradation behavior, multi-stress lifetime modeling has been actively studied with explicit consideration of interaction effects between stresses. Since the 1990s, degradation under combined stresses has been investigated using phenomenological [35], probabilistic [22], and mechanistic approaches [36,37]. Specific studies addressing rotating-machine insulation have also been reported [21, 38]. More recent research has incorporated explicit stress interaction terms [39] and nonlinear stochastic formulations, including Wiener

process-based models [40]. Despite these advances, most existing models are developed based on simplified material systems or assume uniform insulation structures, limiting their applicability to real rotating-machine insulation systems. Several studies have further highlighted the roles of polymer composition, reinforcement structure, and dielectric strength in determining insulation lifetime [28]. This limitation is particularly pronounced for rated-voltage-dependent design variations, where uniform-insulation assumptions restrict generalization across voltage classes [27,29].

2.3. Statistical approaches for ALT/ADT

A major challenge in lifetime modeling arises from sparse, heavily censored, and heterogeneous datasets. To address this issue, modern statistical methodologies have increasingly adopted Bayesian and non-parametric frameworks. Representative examples include Bayesian ALT models with competing risks [9], ADT formulations that incorporate random effects and measurement error [14], and multi-stress ADT reliability analysis for electronic components [15,39,41]. In particular, recent studies have emphasized that degradation-based approaches must explicitly account for multiple sources of uncertainty, including measurement noise, unit-to-unit variability, and latent environmental factors, to obtain reliable lifetime estimates [14,15,42]. Additional studies have compared accelerated lifetime distributions [43], proposed generalized stress-life relationships [44], and demonstrated the methodological advantages of Bayesian inference in reliability analysis [45]. More recently, hybrid approaches combining accelerated testing data with degradation or condition monitoring information have been proposed to mitigate the limitations of failure-based ALT under data scarcity [30,46,47]. Furthermore, uncertainty-aware prognostic frameworks have been developed by integrating probabilistic machine learning and physics-informed models, enabling robust lifetime prediction even in the absence of complete failure labels [32,48,49]. In some cases, auxiliary information derived from diagnostic measurements or predictive models has been utilized to supplement limited failure data; however, such information is inherently indirect and may introduce additional uncertainty if not properly incorporated within a probabilistic framework.

2.4. Research gaps

In summary, three major research gaps can be identified.

- (i) Although combined thermal–electrical stress testing has been standardized, no widely accepted model exists that enables consistent lifetime extrapolation across different insulation designs.
- (ii) Only a limited number of studies explicitly incorporate design covariates, such as R , to evaluate generalization and extrapolation stability across voltage classes.
- (iii) While recent studies have advanced stochastic degradation modeling and uncertainty quantification, the integration of heterogeneous data sources under severe censoring remains challenging.

Motivated by these gaps, the proposed **RIPPLE** framework makes the following contributions:

- Analysis of failure and censored data under simultaneous thermal–electrical stresses to develop lifetime models that more accurately represent actual operating conditions.
- Integration of auxiliary information derived from diagnostic measurements to enhance the effective information content of censored datasets through uncertainty-aware pseudo-label augmentation, treating pseudo-labels as probabilistic soft evidence rather than exact failure observations.
- Proposal and comparative evaluation of a truncated latent-variable likelihood and a variance-inflation likelihood for pseudo-labeled observations, with the former adopted as the primary formulation based on superior empirical performance.

- Formulation of a specification-informed Bayesian AFT model integrating the **R-TNT** lifetime model, which explicitly incorporates rated-voltage as a class-level design covariate enabling consistent lifetime extrapolation across heterogeneous voltage classes, with the additive formulation selected over alternative parameterizations based on Bayesian predictive criteria.

3. Method

Notation. Consider an accelerated life test conducted under m stress combinations, indexed by $i = 1, \dots, m$, each characterized by temperature T_i , applied voltage V_i , and rated-voltage R_i . At each stress level i , a total of n_i specimens are tested, indexed by $j = 1, \dots, n_i$. Over the course of the test, time-series diagnostic measurements $\mathbf{D}_{ij}$ are collected for each specimen (i, j) . The test at stress level i is terminated at time τ_i , by which point n'_i specimens have failed with observed lifetimes $\{y_{ij}\}$; the remaining $c_i = n_i - n'_i$ specimens are right-censored at τ_i .

Framework overview. The proposed **RIPPLE** framework addresses the challenge of lifetime prediction under severe data scarcity by integrating three methodological components within a unified Bayesian inference scheme. First, the **R-TNT** lifetime model extends the standard T-NT formulation by incorporating rated-voltage R as an explicit design covariate, enabling joint modeling across heterogeneous voltage classes (Section 3.2). Second, **pseudo-labeling** generates failure-like lifetime estimates for censored specimens using a deep learning encoder trained on time-series diagnostic measurements and design features (Section 3.4). Third, an uncertainty-aware likelihood treats pseudo-labels as probabilistic soft evidence through a truncated latent-variable formulation, propagating prediction uncertainty into the posterior and enforcing censoring consistency, within a Bayesian inference framework that integrates prior knowledge via MCMC sampling (Sections 3.5–3.6).

3.1. Temperature–Nonthermal (T-NT) lifetime model

In ALT, the scale parameter of a lifetime distribution is commonly modeled using a physically interpretable stress-life relationship that describes how external stresses influence the expected lifetime. When both temperature and a nonthermal stress, such as voltage, affect lifetime, the Arrhenius relationship [4] and the IPL relationship [5,24] can be combined into a unified formulation, commonly referred to as the Temperature–Nonthermal (T-NT) model [50]. A practical form is given by

$$L(T, V) = \theta_1 / V^{-\theta_2} \exp(\theta_3 / T), \quad (1)$$

where L denotes the characteristic lifetime and θ_1 , θ_2 , and θ_3 are model parameters to be estimated. θ_2 governs the inverse power-law dependence of lifetime on voltage stress, and θ_3 is the Arrhenius thermal activation coefficient. The exponential term reflects thermally activated degradation consistent with Arrhenius kinetics, indicating that higher temperature accelerates aging, whereas the power-law term captures the inverse dependence of lifetime on voltage stress.

By taking natural logarithms, the model can be expressed in a log-linear form that is convenient for regression-based estimation:

$$\ln L(T, V) = \ln \theta_1 - \theta_2 \ln V + \theta_3 / T. \quad (2)$$

This log-linear formulation separates thermal and nonthermal effects, thereby enabling joint modeling of temperature and voltage through physically interpretable coefficients.

3.2. Rated-voltage-informed T-NT (R-TNT) lifetime model

Conventional ALT models typically express expected lifetime as a function of temperature and voltage using Arrhenius and

inverse-power relationships [4,5]. In such formulations, the rated-voltage (R) of a machine is often used merely for normalization (e.g., V/R), which implicitly assumes that insulation design remains constant across voltage classes. However, R represents more than a nominal specification: it reflects key design attributes including insulation thickness, conductor spacing, and material composition, which collectively determine the intrinsic robustness of the stator insulation system [27,29]. Empirical studies have shown that insulation systems designed for higher rated-voltages generally exhibit longer lifetimes under equivalent stress conditions due to stronger geometries and higher dielectric strength [17,28].

To account for this inherent design robustness, the Temperature–Nonthermal (T–NT) model is extended by explicitly incorporating R as a covariate that contributes positively to lifetime. The resulting rated-voltage-informed formulation, referred to as the **R–TNT model**, is defined as

$$L(T, V, R) = \frac{\theta_1 \exp(R)}{V^{\theta_2} \exp(-\theta_3/T)}, \quad (3)$$

or equivalently, in logarithmic form,

$$\ln(L) = \ln(\theta_1) + R - \theta_2 \ln(V) + \theta_3/T. \quad (4)$$

The exponential dependence on R captures the strengthening effect associated with insulation design, while the remaining terms describe degradation driven by electrical and thermal stresses. From a failure-mechanism perspective, the influence of R can be interpreted through its relationship with insulation design and electric field strength. For a given applied voltage, insulation systems designed for higher rated-voltage typically employ thicker insulation layers and larger conductor spacing, resulting in lower effective electric field intensity within the dielectric. Since electrical aging mechanisms such as partial discharge initiation, treeing, and insulation degradation are strongly driven by local electric field stress, this reduction in field intensity leads to slower degradation rates and consequently longer lifetimes [17,28].

Although R serves as a practical proxy for insulation design strength, it may not fully capture variations within a given voltage class; therefore, R is treated as a class-level covariate under a fixed kV convention, and any residual heterogeneity and unit dependence are acknowledged as limitations.

3.3. Lifetime distribution

The stress-life models introduced above describe how external stresses (T, V) and design robustness (R) jointly determine the expected lifetime L . In a probabilistic framework, this deterministic stress-life relationship is used to parameterize a statistical lifetime distribution, thereby linking physical stress conditions to the stochastic behavior of failure times.

Let Y denote the lifetime random variable. Two commonly adopted lifetime distributions are considered: the *Weibull* and the *LogNormal* distributions, both of which are widely used in reliability analysis. In each case, the scale parameter α (or $\exp(\mu)$) for the distribution is modeled as a function of (T, V, R) through the R–TNT formulation introduced in Section 3.2.

3.3.1. Weibull distribution

Assume that $Y \sim \text{Weibull}(\alpha, \beta)$, where $\alpha > 0$ and $\beta > 0$ denote the scale and shape parameters, respectively. The corresponding probability density function and survival functions are given by

$$f(y; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{y}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{y}{\alpha}\right)^\beta\right], \quad (5)$$

$$S(y; \alpha, \beta) = \exp\left[-\left(\frac{y}{\alpha}\right)^\beta\right]. \quad (6)$$

Under the R–TNT stress-life formulation, the scale parameter becomes stress-dependent and is expressed as

$$\ln \alpha_i = \ln \theta_1 + R_i - \theta_2 \ln V_i + \theta_3/T_i, \quad (7)$$

which directly replaces the constant scale term in the Weibull distribution. Substituting this form into $f(y; \alpha, \beta)$ embeds the acceleration effects of temperature and voltage, as well as the strengthening effect associated with R , into the probabilistic lifetime model.

3.3.2. LogNormal distribution

Alternatively, lifetime may be modeled using a LogNormal distribution, assuming that the logarithm of lifetime follows a normal distribution,

$$Y \sim \text{LogNormal}(\mu, \sigma^2). \quad (8)$$

In this case, $\mu \in \mathbb{R}$ and $\sigma > 0$ are the log-scale and shape parameters, respectively. The log-scale (location) parameter is linked to (T, V, R) analogously as

$$\mu_i = \ln \theta_1 + R_i - \theta_2 \ln V_i + \theta_3/T_i. \quad (9)$$

This additive log-scale structure is equivalent to an AFT model with R–TNT covariates.

3.3.3. Likelihood formulation under right censoring

Let D denote the observed ALT dataset comprising failure times and censoring information across all m stress levels, and let θ denote the vector of model parameters. The censored likelihood is given by

$$\mathcal{L}(D; \theta) = \prod_{i=1}^m \left[\prod_{j=1}^{n'_i} f(y_{ij}; \theta) \right] [S(\tau_i; \theta)]^{c_i}, \quad (10)$$

where $f(\cdot)$ and $S(\cdot)$ correspond to the Weibull or LogNormal forms defined above. This formulation naturally incorporates right-censored observations through the survival function at the censoring time $S(\tau_i)$.

For insulation systems, however, ALT datasets typically contain a very high proportion of censored observations, resulting in limited information and greater uncertainty in lifetime estimation. This challenge motivates the pseudo-labeling strategy, introduced in Section 3.4, which augments the likelihood with auxiliary lifetime-like targets inferred from diagnostic measurements.

3.4. Pseudo-label generation

Insulation lifetime tests often yield datasets with a high degree of right censoring, which weakens statistical inference and increases predictive uncertainty. To address this limitation, a pseudo-labeling strategy is introduced to estimate lifetime-like targets for censored specimens using diagnostic measurements and design-related features. The proposed pseudo-labeling framework is conceptually related to semi-supervised learning, in which censored specimens are treated as unlabeled samples: their true failure times are unobserved, yet time-series diagnostic indicators together with design parameters encode information about their degradation state. Let y_{ij} denote the true lifetime of specimen j under stress level i , and $\hat{y}_{ij}$ the corresponding pseudo-labeled estimate. The pseudo-labeling model $g^{(s)}$ is trained multiple times with different random seeds, yielding an ensemble of predictions $\{\hat{y}_{ij}^{(s)}\}_{s \in S}$ for each censored specimen, where $S = \{1, \dots, 20\}$ denotes the index set of repeated training runs. The ensemble mean prediction is defined as $\hat{y}_{ij} = \frac{1}{|S|} \sum_{s \in S} \hat{y}_{ij}^{(s)}$.

3.4.1. Input features

Each specimen is characterized by two categories of input features:

- **Time-series features:** $\Delta \tan \delta$, partial discharge magnitude, and insulation resistance measured periodically during the test.
- **Scalar features:** Temperature T , applied stress voltage V , rated-voltage R , and selected design descriptors. Among 42 available design variables, four were retained after expert screening to ensure physical relevance and maintain model interpretability.

All input features were used without interpolation or imputation, as no missing values were present in the dataset. Variable-length sequences were zero-padded to a fixed length equal to the maximum sequence length observed in the dataset, with padding applied at the end of each sequence to preserve the original temporal order. No normalization or scaling was applied to either time-series or scalar features; all inputs were retained in their original physical units.

3.4.2. Pseudo-labeling model architecture

The pseudo-labeling model comprises two parallel branches: a Transformer encoder [51] for time-series diagnostic features and an MLP for scalar features. The resulting latent representations are concatenated and passed to a linear regression head to produce a scalar lifetime prediction $\hat{y}_{ij}$. The architecture is illustrated in Fig. 1.

Time-series branch. The padded sequence $\mathbf{X}_{ij} \in \mathbb{R}^{T_{\max} \times 3}$ is first projected to the model dimension d_{model} via a linear embedding layer, followed by fixed sinusoidal positional encoding to preserve temporal order. The resulting sequence is passed through L stacked Transformer encoder blocks, each comprising multi-head self-attention with $H = 8$ heads and a position-wise feed-forward network (FFN) with hidden dimension $d_{\text{ff}} = 4d_{\text{model}}$, both followed by residual connections and layer normalization.

Scalar branch. The scalar feature vector contains seven variables: test temperature T , applied voltage V , rated-voltage R , and four expert-selected design descriptors. This input is processed by a two-layer MLP:

$$\mathbb{R}^7 \rightarrow \mathbb{R}^{d_{\text{model}}/2} \rightarrow \mathbb{R}^{d_{\text{model}}},$$

with ReLU activation after the first linear layer. The resulting scalar representation is denoted by $\mathbf{z}_{ij}^{\text{sc}} \in \mathbb{R}^{d_{\text{model}}}$.

Fusion and regression. The two latent representations are concatenated,

$$\mathbf{z}_{ij} = [\mathbf{z}_{ij}^{\text{ts}} \parallel \mathbf{z}_{ij}^{\text{sc}}] \in \mathbb{R}^{2d_{\text{model}}},$$

and passed to a linear regression head to predict the pseudo-labeled lifetime:

$$\hat{y}_{ij} = \mathbf{w}^T \mathbf{z}_{ij} + b.$$

The overall model achieves approximately 19× faster training than the standard Transformer base model [51] while maintaining comparable representational capacity; full computational benchmarks are provided in the Supplementary Material (S2).

3.4.3. Training

To reduce the risk of lifetime underestimation, an asymmetric squared-error loss function is adopted:

$$\ell_{ij} = \begin{cases} \gamma(y_{ij} - \hat{y}_{ij})^2, & \text{if } y_{ij} > \hat{y}_{ij}, \\ (y_{ij} - \hat{y}_{ij})^2, & \text{otherwise,} \end{cases} \quad \ell = \frac{1}{N} \sum_{i,j} \ell_{ij}, \quad (11)$$

where $\gamma > 1$ increases the penalty associated with early, potentially unsafe predictions. Each model in the ensemble is trained independently, and the loss is defined with respect to the prediction from an individual model. Pseudo-labeled lifetimes $\hat{y}_{ij}$ are subsequently generated for censored specimens using the trained ensemble.

Gradient-based saliency mapping [52] is used to quantify the contribution of individual features to lifetime prediction, averaged across ensemble members. The resulting saliency patterns are interpreted with expert input to verify their consistency with established degradation mechanisms, such as the dominant influence of temperature over voltage. This analysis also evaluates the contribution of R relative to other design descriptors, thereby confirming its suitability as a proxy for insulation robustness.

3.5. Uncertainty-aware likelihood

Since true failure times are unavailable for censored specimens, specimen-specific aleatoric uncertainty cannot be directly estimated. Instead, the log-scale residual variance computed on validation data,

$$\sigma_{\text{resid}}^2 = \text{Var}(\log y^{\text{val}} - \log \hat{y}^{\text{val}}), \quad (12)$$

is used as a global residual uncertainty. In addition, the dispersion of ensemble predictions provides a sample-specific uncertainty estimate,

$$\sigma_{\text{seed},ij}^2 = \text{Var}(\log \hat{y}_{ij}^{(s)}). \quad (13)$$

These two components are combined to define the pseudo-label uncertainty for each censored specimen as

$$s_{ij}^2 = \sigma_{\text{resid}}^2 + \sigma_{\text{seed},ij}^2. \quad (14)$$

This uncertainty quantification allows pseudo-labels to be treated as probabilistic, rather than deterministic, estimates of lifetime. These quantities are incorporated into the likelihood function to enable uncertainty-aware inference under heavy censoring.

With μ_i as given in Eq. (9), two formulations are introduced.

(1) Variance-inflation formulation: In the LogNormal model, pseudo-labels are treated as noisy observations of the latent mean,

$$\ln \hat{y}_{ij} \sim \mathcal{N}(\mu_i, \sigma^2 + s_{ij}^2), \quad (15)$$

where μ_i is given by Eq. (9), σ^2 is the model variance parameter, and s_{ij}^2 is the pseudo-label uncertainty. This formulation increases the effective variance for uncertain pseudo-labels, thereby reducing their influence on parameter estimation.

(2) Latent-variable formulation: A more principled alternative introduces a latent true log-lifetime $u_{ij} = \ln Y_{ij}$ for each censored specimen,

$$u_{ij} \sim \mathcal{N}(\mu_i, \sigma^2), \quad u_{ij} > \ln \tau_i, \quad (16)$$

$$\ln \hat{y}_{ij} \sim \mathcal{N}(u_{ij}, s_{ij}^2), \quad (17)$$

where the truncation enforces consistency with the censoring constraint. In this formulation, pseudo-labels are interpreted as noisy measurements of the latent failure time. This hierarchical structure separates the underlying lifetime generation process from the pseudo-label observation process, allowing uncertainty to be propagated through the latent variable. The truncated latent-variable formulation is adopted as the primary approach based on its superior empirical performance (Section 4.4.2).

Weibull extension: For the Weibull model, the additive variance structure used in the LogNormal case is not directly applicable. Instead, pseudo-label uncertainty is incorporated by downweighting their likelihood contributions. Specifically, a weight

$$w_{ij} = \frac{1}{1 + s_{ij}^2/c} \quad (18)$$

is applied to the log-likelihood term of each pseudo-labeled observation, where c is a fixed scaling constant. This ensures that pseudo-labels with higher uncertainty contribute less to parameter estimation.

3.6. Bayesian inference framework

Model parameters are estimated within a Bayesian inference framework. For the Weibull formulation, the parameter vector is $\theta = (\theta_1, \theta_2, \theta_3, \beta)$, whereas for the LogNormal formulation it is $\theta = (\theta_1, \theta_2, \theta_3, \sigma^2)$. This framework enables the integration of prior knowledge with experimental data while providing comprehensive uncertainty quantification.

3.6.1. Prior distributions

Prior distributions are selected to reflect physically meaningful parameter ranges while maintaining sufficient flexibility. Weakly informative priors (e.g., Gamma, Normal, or Uniform families) are assigned to scale and stress coefficients to prevent unrealistic estimates without imposing overly restrictive assumptions [45,53]. These priors function as soft regularizers, which are particularly beneficial in small-sample settings characterized by heavy censoring. Numerical details of the prior specifications are presented in Section 4.3.

3.6.2. Posterior inference

Given the likelihood function $\mathcal{L}(D | \theta)$ defined in Eq. (10) and a prior distribution $p(\theta)$, Bayes' theorem yields the posterior distribution

$$p(\theta | D) \propto p(\theta) \mathcal{L}(D | \theta), \quad (19)$$

where D denotes the observed dataset, including both failure times and right-censored observations. The resulting posterior distribution characterizes the full uncertainty associated with the model parameters and enables the estimation of credible intervals, predictive distributions, and reliability metrics [45,53,54].

For the latent-variable formulation introduced in Eq. (16), inference is performed over both the model parameters and the latent log-lifetimes of censored specimens. Let $\mathbf{u} = \{u_{ij}\}$ denote the collection of latent variables. The corresponding joint posterior is

$$p(\theta, \mathbf{u} | D) \propto p(\theta) p(\mathbf{u} | \theta) p(D | \mathbf{u}, \theta), \quad (20)$$

where $p(\mathbf{u} | \theta)$ represents the truncated latent lifetime model and $p(D | \mathbf{u}, \theta)$ represents the observation model linking pseudo-labels to latent lifetimes. MCMC sampling proceeds jointly over θ and $\mathbf{u}$, propagating pseudo-label uncertainty through the posterior and yielding statistically coherent uncertainty quantification.

Since the normalizing constant is generally intractable, posterior samples are obtained using Markov chain Monte Carlo (MCMC) sampling via the No-U-Turn Sampler (NUTS) [55]. Convergence and sampling quality are assessed using standard diagnostics, including the potential scale reduction factor, effective sample size, and visual inspection of trace plots.

The posterior predictive lifetime under new operating conditions (T^*, V^*, R^*) is obtained as

$$p(y^* | D, T^*, V^*, R^*) = \int p(y^* | \theta, T^*, V^*, R^*) p(\theta | D) d\theta. \quad (21)$$

For the latent-variable formulation, this predictive distribution is induced from the marginal posterior of θ after integrating over the latent variables. This predictive distribution supports direct reliability estimation and enables extrapolation across voltage classes under previously untested conditions.

The complete RIPPLE workflow, integrating pseudo-label generation with Bayesian R-TNT inference, is summarized in Algorithm 1.

Algorithm 1: RIPPLE Workflow

Require:

Stress levels $i = 1, \dots, m$ with (T_i, V_i, R_i, τ_i)
 Failures $\{y_{ij}\}_{j=1}^{n'_i}$, Censored specimens $\{y_{ij}\}_{j=n'_i+1}^{n_i}$
 Diagnostic features $\mathbf{D}_{ij}$
 Prior $p(\theta)$

1. Train ensemble pseudo-labeler

for $s \in S$ do

Train model $g^{(s)}$ using $(\mathbf{D}_{ij}, T_i, V_i, R_i) \rightarrow y_{ij}$ for $j \leq n'_i$

end for

2. Generate pseudo-labels and uncertainty

for $i = 1$ to m do

for $j = n'_i + 1$ to n_i do

Compute $\hat{y}_{ij}^{(s)} = g^{(s)}(\mathbf{D}_{ij}, T_i, V_i, R_i)$ for all s

Compute ensemble mean $\hat{y}_{ij} = \frac{1}{|S|} \sum_{s \in S} \hat{y}_{ij}^{(s)}$

Compute uncertainty $s_{ij}^2 = \text{Var}(\ln \hat{y}_{ij}^{(s)}) + \sigma_{\text{resid}}^2$

end for

end for

3. Construct uncertainty-aware likelihood

Form likelihood using observed failures and pseudo-labels
 Select formulation:

(a) Variance-inflation: $\ln \hat{y}_{ij} \sim \mathcal{N}(\mu_i, \sigma^2 + s_{ij}^2)$

(b) Latent-variable: introduce u_{ij} with $u_{ij} > \ln \tau_i$

4. Bayesian inference with joint sampling

Define posterior $p(\theta, \mathbf{u} | D)$ (if latent-variable formulation is used)

Perform MCMC sampling over θ (and $\mathbf{u}$ if applicable)

5. Prediction

Compute posterior predictive $p(y^* | T^*, V^*, R^*, D)$

4. Experiments

4.1. Data

The dataset was obtained from ALT conducted on stator insulation systems used in high-voltage rotating machines. Each specimen consisted of a form-wound coil segment designed for either the 6.6 kV or 13.8 kV rated-voltage class. The primary objective of the experiments was to evaluate insulation lifetime under combined thermal-electrical stresses and to collect auxiliary diagnostic measurements for subsequent pseudo-label estimation. Tables 1 and 2 summarize the applied stress conditions, test outcomes, and observed failure times.

4.1.1. Insulation system description

The insulation system of a rotating-machine protects conductors from electrical, thermal, and mechanical stresses encountered during operation. It comprises multiple dielectric layers that form the turn, slot, and end-winding insulation, which collectively provide long-term resistance to localized failure and environmental degradation.

Table 1
Applied stress levels, failure counts, and censoring times.

Rated volt.	Level	Volt. (kV)	Temp. (°C)	Failures	Censoring time (h)
13.8 kV	Level 1	23.46	145	2 out of 5	6000
	Level 2	24.84	150	3 out of 5	1152
	Level 3	26.22	155	2 out of 5	604.17
6.6 kV	Level 1	11.22	145	3 out of 5	6561
	Level 2	12.54	155	3 out of 6	716
	Level 3	13.86	165	2 out of 6	528

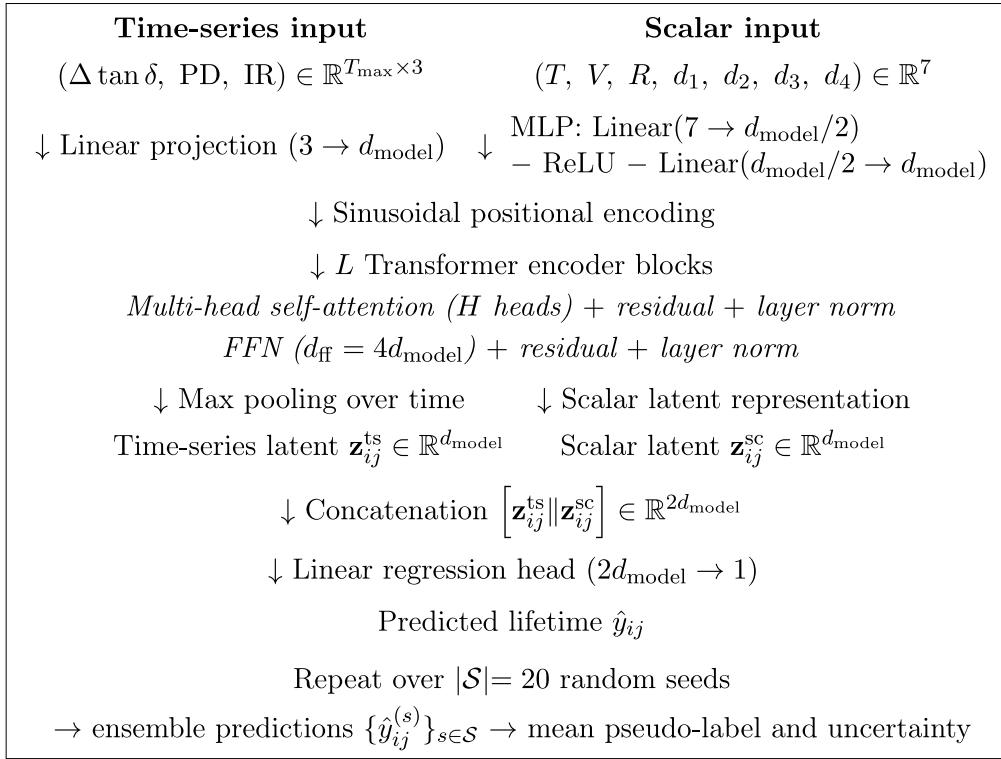


Fig. 1. Schematic of the pseudo-labeling Transformer–MLP architecture. A Transformer encoder extracts temporal degradation features from diagnostic sequences, while an MLP processes scalar covariates. Their latent representations are fused and mapped to a scalar lifetime prediction. Multiple independently trained models are ensemble to quantify epistemic uncertainty.

Table 2

Observed failure times for stator insulation specimens.

Rated volt.	Stress level	Failure time (h)
13.8 kV	Level 1	4435, 4559
	Level 2	921, 1130, 1075
	Level 3	604, 447
6.6 kV	Level 1	6348, 5445, 6108
	Level 2	398, 627, 401
	Level 3	457, 434

4.1.2. Test procedure

ALTs were conducted in accordance with IEC 60034-18-33 [20], which specifies combined thermal–electrical endurance procedures for form-wound coils. All test specimens were manufactured using actual production processes to ensure that the insulation structures were representative of in-service conditions.

Each aging sub-cycle consisted of thermal aging followed sequentially by mechanical vibration and moisture exposure, and then diagnostic and voltage testing. Mechanical testing applied vibration amplitudes of 0.2–0.3 mm at frequencies of 50–60 Hz for one hour (corresponding to acceleration levels of approximately 1.5–1.8 g), while moisture exposure tests subjected specimens to conditions of 25 °C and 95% relative humidity for a minimum duration of 48 h. All diagnostic and voltage tests were performed immediately thereafter, while visible moisture remained on the coil surfaces.

The aging sub-cycles were repeated approximately ten times. This resulted in a total test duration of 1 year and 2 months for the 6.6 kV specimens (2017.03.14 to 2018.05.08) and 1 year and 1 month for 13.8 kV specimens (2020.11.03 to 2021.12.16). The duration of individual sub-cycles varied between 2 and 25 days depending on the applied stress level.

4.1.3. Diagnostic indicators

Three diagnostic indicators were monitored at each aging sub-cycle to characterize insulation degradation:

- $\Delta \tan \delta$: The difference in dielectric loss tangent measured at low voltage (20% of rated) and at the rated-voltage conditions. This indicator typically increases with insulation aging, moisture ingress, or surface contamination, and therefore serves as a sensitive indicator of changes in dielectric polarization behavior.
- Partial discharge (PD) magnitude: The apparent charge associated with localized discharges within or on the surface of the insulation system under high voltage stress. PD activity reflects the development of internal voids or surface defects and provides indirect evidence of progressive loss of insulation endurance across sub-cycles.
- Insulation resistance (IR): The direct-current resistance measured between conductors and ground, expressed in M Ω or G Ω . A reduction in IR indicates the formation of additional leakage current paths due to insulation aging or moisture absorption.

4.1.4. Data collection procedure

All tests were conducted in environmental chambers designed to maintain uniform thermal conditions while allowing safe application of high-voltage. The diagnostic indicators $\Delta \tan \delta$, PD magnitude, and IR were measured following each aging sub-cycle as part of a sequence of conditioning and diagnostic tests performed on surviving specimens. After completion of each sub-cycle, specimens were subjected to conditioning procedures, including mechanical and moisture conditioning, followed by voltage and other diagnostic tests in accordance with standard test protocols. Individual sub-cycles lasted from several days to several weeks, depending on the applied stress level, and terminated either in dielectric breakdown or in right-censoring at the end of the observation window. Failures were defined as breakdown events occurring in either the turn insulation or the mainwall insulation,



Fig. 2. Overview of the ALT setup. Left: diagnostic configuration for partial discharge, dielectric loss tangent ($\tan \delta$), and insulation resistance measurements; Right: high-temperature and high-voltage endurance chamber and transformer-based power supply system.

Table 3

Hyperparameter search space used for pseudo-labeling model training.

Hyperparameter	Search space
Encoder blocks	{6, 5, 4}
Hidden dimensions	{64, 128, 256, 512, 1024}
Asymmetry weight γ	{2, 1.5, 1.2, 1.0}
Learning rate	$\{10^{-3}, 5 \times 10^{-4}, 10^{-4}\}$
Weight decay	{1e-2, 1e-3, 1e-4}

whereas specimens that did not fail within the observation window were treated as right-censored. The resulting dataset comprises observed failure times, censoring times, and time-series measurements of diagnostic indicators, and it supports both pseudo-label generation and Bayesian reliability modeling. The overall test environment and diagnostic configuration employed for the combined thermal-electrical ALT are illustrated in Fig. 2.

Data collection was subject to practical constraints including limited chamber capacity, bounded stress levels to preserve representative aging mechanisms, and high specimen fabrication costs; full details are provided in Supplementary Material (S1).

4.2. Pseudo-labeling results

To address the limited number of observed failures, pseudo-labeled lifetimes were generated using the framework described in Section 3.4. A comprehensive hyperparameter search was performed over key model settings, including encoder depth, hidden dimension size, loss penalty weight γ , learning rate (LR), and weight decay (WD), as summarized in Table 3. Model training was conducted exclusively on specimens with observed failures, with a 60:40 train-validation split stratified by stress condition. Censored specimens were excluded from the training process and were assigned pseudo-labeled lifetimes only during the inference stage.

All hyperparameter configurations are evaluated based on the aggregated ensemble predictions, ensuring that model selection reflects the behavior of the final pseudo-labeling scheme rather than individual stochastic realizations.

Model selection was guided by four criteria: (i) predictive accuracy on validation data; (ii) consistency with censoring constraints; (iii) alignment of feature importance with physical intuition; and (iv) avoidance of excessive lifetime underestimation.

4.2.1. Hyperparameter tuning results

Table 4 summarizes the top-performing hyperparameter configurations. The configurations are ranked based on validation MAPE computed from ensemble predictions, and the table reports their corresponding MSE and underestimation rates (UE (%)). As shown, the configuration with $\gamma = 1.0$ achieves the lowest validation MAPE but the

Table 4

Top-5 hyperparameter configurations ranked by validation MAPE. Both MAPE and MSE are computed on validation samples with observed lifetimes, since censored samples do not have true lifetime labels. LR and WD denote learning rate and weight decay, respectively. UE denotes the underestimation rate on validation samples. Bold denotes the best value in each column.

Blocks	Hidden	γ	LR	WD	MAPE	MSE	UE (%)
6	1024	1.0	10^{-3}	10^{-3}	0.178	63 781	42.9
6	1024	1.2	10^{-3}	10^{-4}	0.185	58 713	28.6
4	1024	1.2	10^{-3}	10^{-4}	0.188	52 585	23.2
4	1024	1.2	10^{-3}	10^{-3}	0.189	52 162	31.5
6	1024	1.0	10^{-3}	10^{-4}	0.183	66 813	40.7

Table 5

Sensitivity analysis of the asymmetry parameter γ . For each value of γ , Mean MAPE denotes the average predictive error across all configurations sharing that value, and Best MAPE corresponds to the lowest error achieved within the group. UE denotes the underestimation rate and VR the violation rate, both evaluated on the representative configuration (lowest validation MAPE) within each group. UE is computed on validation samples with observed lifetimes; VR is computed on censored samples and quantifies the proportion of predictions falling below the censoring time.

γ	Mean MAPE	Best MAPE	UE (%)	VR (%)
1.0	0.323	0.178	42.9	26.7
1.2	0.332	0.185	28.6	20.0
1.5	0.348	0.196	57.1	40.0
2.0	0.368	0.272	42.9	26.7

highest underestimation rate (42.9%). Since conservative predictions are preferred in reliability applications, the configuration with Blocks = 4, Hidden = 1024, $\gamma = 1.2$, LR = 10^{-3} , and WD = 10^{-4} (Row 3) was selected based on its favorable balance between predictive accuracy and underestimation risk.

4.2.2. Sensitivity to the asymmetry parameter γ

Table 5 summarizes the sensitivity of model performance to γ . The relationship between γ and UE/VR is non-monotonic: UE decreases from 42.9% at $\gamma = 1.0$ to a minimum of 28.6% at $\gamma = 1.2$, but then rises sharply to 57.1% at $\gamma = 1.5$ before partially recovering at $\gamma = 2.0$. This non-monotonic pattern is attributed to the interaction between penalty strength and optimization landscape: at moderate $\gamma = 1.2$, the asymmetric loss steers the ensemble toward conservative predictions without disrupting convergence, whereas $\gamma \geq 1.5$ introduces optimization instability that paradoxically elevates underestimation. Taken together, $\gamma = 1.2$ achieves the lowest UE (28.6%) and VR (20.0%) while incurring only a marginal MAPE increase of 0.007 relative to $\gamma = 1.0$, confirming that the asymmetric loss is most effective at a moderate level.

Table 6

Representative validation results for specimens with observed failures.

Sample	Predicted lifetime	Observed lifetime
1	635	457
2	533	540
3	4510	4435
4	925	921

Table 7

Representative pseudo-labeled lifetime predictions for censored specimens.

Sample	Predicted lifetime	Censoring time
1	6580	6000
2	6280	6000
3	464	528
4	6713	6561

4.2.3. Model validation and physical consistency

To further validate the selected pseudo-labeling model, we examine its predictive behavior on both observed failure samples and censored specimens. For specimens with observed failures, Table 6 shows that the predicted lifetimes are in close agreement with the observed values, indicating that the model successfully captures the dominant degradation dynamics.

For censored specimens, Table 7 demonstrates that the predicted lifetimes generally exceed the corresponding censoring times, satisfying the physical constraint $\hat{y}_i \geq \tau_i$. Few exceptions are noted but handled probabilistically in the subsequent Bayesian modeling stage and treated as noisy labels with uncertainty.

In addition, feature attribution analysis based on gradient-based saliency further confirmed that temperature has the strongest influence on lifetime prediction, followed by applied voltage, while R plays a significant role among design-related factors. This alignment with domain knowledge provides further support for the validity of the selected pseudo-labeling.

4.3. Bayesian inference and model comparison

4.3.1. Prior specification

Both diffuse and informative prior distributions were evaluated to assess their influence on model stability and inference robustness. Diffuse priors (Uniform1 & Uniform2, Gamma1) imposed minimal constraints on parameter estimates, thereby allowing the data to dominate posterior inference. In contrast, informative priors (Uniform3, Gamma2) were parameterized using Arrhenius and inverse-power law coefficients estimated from single-stress ALTs of the 6.6 kV insulation system. Specifically, the Arrhenius temperature coefficient was centered near 17,440 based on thermal-only ALT data, while the inverse-power law coefficient was centered near 16.4312 based on voltage-only ALT data. Although single-stress experiments were conducted for both the 6.6 kV and 13.8 kV classes, only the 6.6 kV dataset yielded a sufficient number of failures across multiple thermal levels to support stable parameter estimation. Under the assumption that the temperature acceleration parameter is invariant across voltage classes, the priors derived from the 6.6 kV dataset were therefore adopted as physically grounded informative priors for all analyses. As the Weibull and LogNormal lifetime formulations differ in parameterization (scale vs. log-scale), prior ranges were adjusted to ensure consistency across distributions. The same prior specifications were used across all likelihood formulations to ensure a fair comparison of modeling choices. Complete prior specifications for both models are provided in Tables 8 and 9.

Table 8

Prior specifications for the Weibull lifetime distribution.

Prior	θ_1	θ_2	θ_3	β
Uniform1	$\mathcal{U}(0, 25)$	$\mathcal{U}(0, 25)$	$\mathcal{U}(0, 25)$	$\mathcal{U}(0, 1)$
Uniform2	$\mathcal{U}(0.0100)$	$\mathcal{U}(0.0100)$	$\mathcal{U}(0.0100)$	$\mathcal{U}(0, 1)$
Uniform3	$\mathcal{U}(0.0100)$	$\mathcal{U}(0, 32.8624)$	$\mathcal{U}(0, 34880)$	$\mathcal{U}(0, 1)$
Gamma1	$\Gamma(1, 1)$	$\Gamma(1, 1)$	$\Gamma(1, 1)$	$\Gamma(1, 1)$
Gamma2	$\Gamma(1, 1)$	$\Gamma(1, 0.06086)$	$\Gamma(5, 0.0002867)$	$\Gamma(1, 1)$

Table 9

Prior specifications for the LogNormal lifetime distribution.

Prior	θ_1	θ_2	θ_3	σ^2
Uniform1	$\mathcal{U}(0.0100)$	$\mathcal{U}(0.0100)$	$\mathcal{U}(0.0100)$	$\mathcal{U}(0, 1)$
Uniform2	$\mathcal{U}(0, 10000)$	$\mathcal{U}(0, 10000)$	$\mathcal{U}(0, 10000)$	$\mathcal{U}(0, 1)$
Uniform3	$\mathcal{U}(0, 10000)$	$\mathcal{U}(0, 32.8624)$	$\mathcal{U}(0, 34880)$	$\mathcal{U}(0, 1)$
Gamma1	$\Gamma(1, 0.01)$	$\Gamma(1, 0.01)$	$\Gamma(1, 0.01)$	$\Gamma(1, 0.01)$
Gamma2	$\Gamma(1, 0.01)$	$\Gamma(1, 0.06086)$	$\Gamma(5, 0.0002867)$	$\Gamma(1, 0.1)$

4.3.2. Bayesian implementation

Posterior inference was performed using PyMC [56] with the NUTS Sampler [55]. For the uncertainty-aware likelihood formulations, posterior sampling was performed under both the variance-inflation (Eq. (15)) and latent-variable (Eq. (16)) formulations. In the latent-variable formulation, inference was carried out jointly over model parameters and latent log-lifetimes of censored specimens, resulting in joint posterior sampling of θ and $\mathbf{u}$. For the Weibull extension, the scaling constant c is set to the posterior mean of σ^2 from the LogNormal model. Each Markov chain was executed with 5000 warm-up iterations and 1000 posterior draws after burn-in, using a target acceptance rate of 0.9. Standard convergence diagnostics indicated satisfactory chain mixing and stable posterior estimates across all runs.

4.3.3. Evaluation metrics and results

Model fit is assessed using WAIC (Widely Applicable Information Criterion) and PSIS-LOO (Pareto-Smoothed Importance Sampling Leave-One-Out cross-validation), which estimate out-of-sample predictive accuracy from the full posterior; the maximum Pareto $\hat{k}$ serves as a reliability diagnostic for the LOO approximation, with values below 0.7 indicating trustworthy estimates [57]. Point-prediction accuracy is assessed via MAPE (Mean Absolute Percentage Error), R^2 (coefficient of determination), and MPE (Mean Percentage Error), all computed exclusively on specimens with observed failures. MPE is defined as $\frac{1}{N} \sum_{i,j} \frac{y_{ij} - \hat{y}_{ij}}{y_{ij}}$, and serves as a signed metric that distinguishes systematic overestimation (negative values) from underestimation (positive values), with values close to zero indicating negligible directional bias.

Incremental contribution and overall model performance. Table 10 presents a step-by-step evaluation from the censored-likelihood baseline (Eq. (10)) to RIPPLE (Eq. (16)), enabling both component-level interpretation and overall performance comparison. RIPPLE achieves substantial gains over the censored-likelihood baseline: MAPE improves from 2.213 to 0.453 and R^2 from 0.000 to 0.885, with each component contributing cumulatively. Pseudo-labeling provides the dominant effect, while the uncertainty-aware likelihood and latent-variable truncation offer complementary improvements in robustness and statistical consistency. The maximum Pareto $\hat{k}$ remains stable across all steps, confirming that the LOO estimates are reliable.

Sensitivity to prior specification. Table 11 evaluates sensitivity to prior specification. Among the well-specified configurations (Uniform2, Uniform3, and Gamma2), performance differences are modest ($\Delta\text{WAIC} \leq 0.26$), demonstrating robustness to prior choice within a physically reasonable range. Gamma2 achieves the best predictive accuracy (MAPE and R^2), reflecting the benefit of incorporating physical knowledge into the prior. The severe degradation observed for Gamma1 and Uniform1 is attributable to their prior bounds being inconsistent with the physically expected parameter magnitudes rather than to diffuseness per se.

Table 10

Incremental contribution of each modeling component from the baseline (Step 1) to RIPPLE (Step 4). Step 1 (baseline): R-TNT model with censored likelihood only (observed failures and right-censored observations; no pseudo-labels). Step 2 adds deterministic pseudo-labeling (pseudo-labels treated as exact observations). Step 3 introduces an uncertainty-aware likelihood via the variance-inflation formulation, treating pseudo-labels as noisy observations. Step 4 incorporates the latent-variable truncated likelihood formulation. ↓ lower is better; ↑ higher is better. Bold denotes the best value in each column.

Change	WAIC ↓	LOO ↓	MAPE ↓	MPE	R^2 ↑	max Pareto $\hat{k}$
Censored-likelihood baseline	283.405	283.434	2.213	−2.213	0.000	0.335
+ deterministic pseudo-labeling	270.259	270.281	0.862	−0.776	0.849	0.232
+ uncertainty-aware likelihood	269.567	269.589	0.802	−0.708	0.896	0.264
+ latent-variable truncation	268.412	268.505	0.453	−0.176	0.885	0.331

Table 11

Prior sensitivity analysis under the truncation-based formulation. Differences (Δ) are computed relative to the best-performing prior.

Prior	WAIC ↓	LOO ↓	MAPE ↓	MPE	R^2 ↑	max Pareto $\hat{k}$
Uniform1	411.017	411.019	5.191	−413.571	0.000	0.388
Uniform2	268.155	268.229	0.480	−18.551	0.839	0.366
Uniform3	268.329	268.397	0.476	−17.722	0.833	0.319
Gamma1	328.137	328.157	20.156	−1966.519	0.000	0.255
Gamma2	268.412	268.505	0.453	−17.624	0.885	0.331

Table 12

Posterior summaries for the final model (R-TNT LogNormal with Gamma2 prior).

Parameter	Mean	SD	2.5% HDI	97.5% HDI
θ_1	99.03	96.6	0.19	292.76
θ_2	10.38	0.33	9.78	11.06
θ_3	9525.78	643.32	8269.82	10757.79
σ^2	0.24	0.07	0.12	0.38

Posterior interpretation. Table 12 presents posterior summaries of the final model (R-TNT LogNormal with Gamma2 prior). The stress coefficients θ_2 and θ_3 are well-identified (CV: 3.2% and 6.7%, respectively), with $\hat{\theta}_2 = 10.38$ (95% HDI: [9.78, 11.06]) and $\hat{\theta}_3 = 9526$ K (95% HDI: [8,270, 10,758] K), corresponding to an estimated activation energy of $\hat{E}_a \approx 0.821$ eV. By contrast, the pre-exponential intercept θ_1 exhibits a wide marginal posterior (CV = 97.5%), reflecting the collinearity between $\ln \theta_1$ and the R covariate: the two voltage classes (6.6 and 13.8 kV) induce an additive shift of $\Delta R = 7.2$ in the log-lifetime predictor, which is partially exchangeable with $\ln \theta_1$ at the marginal level. The Gamma2 prior regularizes θ_1 toward physically plausible values, and the predictive distribution remains well-calibrated regardless. This identifiability limitation of θ_1 is acknowledged as a direction for methodological refinement when more than two voltage classes become available. Alternative parameterizations of R —including log-additive and scaled variants—were evaluated in the ablation study (Section 4.4.5) but consistently underperformed the direct-additive formulation across all Bayesian predictive criteria; the raw-kV convention is therefore retained, with unit dependence acknowledged as a limitation.

4.3.4. Predictive reliability and uncertainty reduction

To evaluate the impact of pseudo-labeling on reliability estimation, predictive reliability was assessed at $T = 155^\circ\text{C}$ and $V = 2.1R$ using posterior samples from the LogNormal model with Gamma2 prior. Fig. 3 shows that the baseline reliability curves lie consistently above those of RIPPLE, and this observation is corroborated by the severe point estimation errors against observed failures reported in Table 10: the baseline yields MPE of −221.276, MAPE of 2.213, and R^2 of 0.000, confirming systematic overestimation of reliability. Subsequent steps in Table 10 show a monotonic reduction in directional bias (MPE: −221.276 → −77.631 → −70.825 → −17.624), with both the deterministic pseudo-labeling model and RIPPLE producing progressively more conservative reliability estimates.

Table 13 quantifies the associated predictive uncertainty using the width of the 90% Bayesian credible interval (BCI). RIPPLE achieves approximately 90% reduction in BCI width compared to the censored-likelihood baseline for both voltage classes (13.8 kV: 145.21 vs. 1425.09; 6.6 kV: 126.79 vs. 1382.58), substantially outperforming the deterministic pseudo-labeling model. Overall, RIPPLE produces sharper predictive distributions with substantially improved point-prediction accuracy, while maintaining reasonable though sub-nominal coverage.

4.3.5. Posterior predictive check and distributional comparison

To evaluate model adequacy beyond reliability curves, posterior predictive checks (PPC) were conducted for the latent-variable formulation (RIPPLE), the variance-inflation formulation, and the Weibull extension. Fig. 4 presents both observation-level and pooled PPC results. The variance-inflation formulation and the Weibull extension exhibit very wide credible intervals that cover nearly all observed failures, reflecting over-dispersion rather than accurate modeling. In contrast, RIPPLE produces substantially tighter credible intervals while still capturing most observed failures.

Table 14 further quantifies this trade-off. Although RIPPLE achieves slightly lower coverage (0.883) compared to the other models, it yields significantly smaller median absolute error (288.1 vs. 462.9 for variance-inflation and 935.2 for the Weibull extension). The high coverage observed in the variance-inflation formulation and the Weibull extension is primarily driven by their excessively wide credible intervals, which can mask poor predictive accuracy—highlighting the limitation of coverage as a standalone metric. These results indicate that the latent-variable formulation achieves a better balance between calibration and sharpness, producing more informative and practically useful predictive distributions compared to alternative approaches.

4.4. Ablations

4.4.1. T-NT vs. R-TNT

Table 15 compares cross-class predictive performance between the conventional T-NT formulation and the R-TNT model. When trained on one voltage class and evaluated on another, the T-NT model exhibits substantial degradation, reflected in significantly lower ELPD values. In contrast, the R-TNT model maintains strong predictive performance across voltage classes, achieving large improvements in both ELPD and average log predictive density. This result demonstrates that explicitly incorporating rated-voltage as a proxy for design strength enables the model to capture cross-class variability and substantially improves generalization.

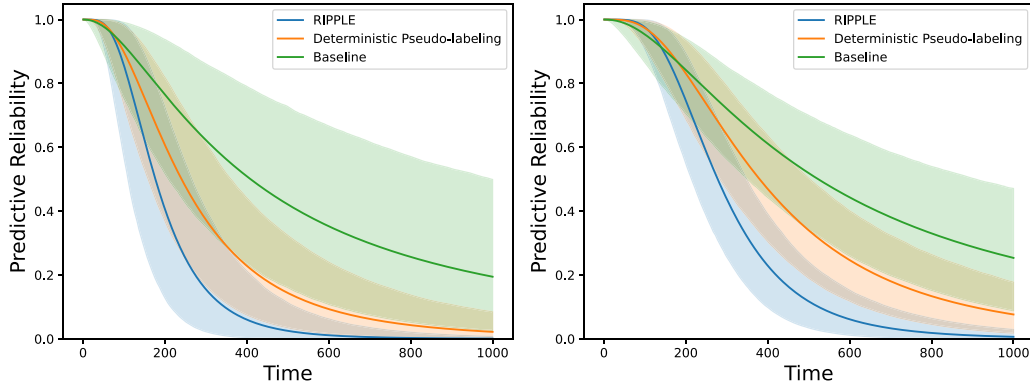


Fig. 3. Predictive reliability curves at $T = 155^\circ\text{C}$ and $V = 2.1R$ for the 13.8 kV (left) and 6.6 kV (right) classes. RIPPLE produces more conservative and better-calibrated reliability estimates compared to the baseline and pseudo-label-only models. Shaded regions represent 90% Bayesian credible intervals (BCIs).

Table 13

Comparison of predictive uncertainty based on the width of the 90% credible interval (BCI).
Reduction is computed relative to the censored-likelihood baseline model.

Voltage	Model	BCI width	Reduction vs. baseline (%)
13.8 kV	RIPPLE	145.21	89.81
13.8 kV	Deterministic pseudo-labeling	224.84	84.22
13.8 kV	Censored-likelihood baseline	1425.09	0.00
6.6 kV	RIPPLE	126.79	90.83
6.6 kV	Deterministic pseudo-labeling	255.98	81.49
6.6 kV	Censored-likelihood baseline	1382.58	0.00

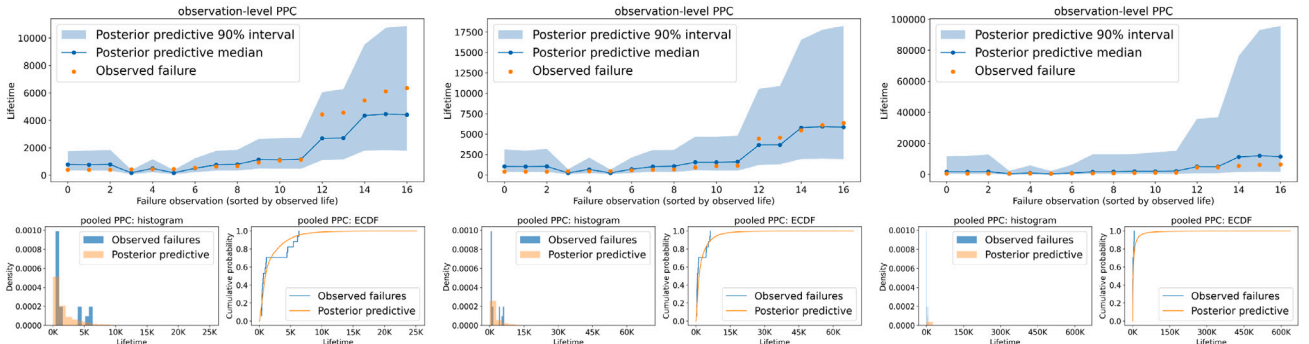


Fig. 4. Posterior predictive checks (PPC) for three modeling approaches: the latent-variable formulation (RIPPLE), the variance-inflation formulation, and the Weibull extension. The top row shows observation-level PPC with posterior predictive medians and 90% credible intervals, while the bottom row presents pooled distribution comparisons using histogram and ECDF.

Table 14

Posterior predictive performance comparison across models. Coverage₉₀ denotes the proportion of observed failures contained within the 90% posterior predictive interval, and the median absolute error is computed with respect to the posterior predictive median.

Model	Coverage ₉₀	Median Abs. Error
Latent-variable formulation (RIPPLE)	0.883	288.115
Variance-inflation formulation	1.00	462.866
Weibull extension	1.00	935.234

The Δ ELPD gain is substantially larger in the $6.6 \rightarrow 13.8$ direction (1287.940) than in the $13.8 \rightarrow 6.6$ direction (92.564). This asymmetry reflects the role of the Gamma2 prior, which was derived from 6.6 kV data: in the $6.6 \rightarrow 13.8$ setting, neither the training data nor the prior provides information about the 13.8 kV class, leaving the T-NT model severely underconstrained. In the reverse direction, the 6.6 kV-informed prior partially compensates for the absence of 6.6 kV training data, reducing the baseline degradation and narrowing the margin for R-TNT to improve upon.

4.4.2. Pseudo-label constraint handling

To isolate the effect of pseudo-label constraint handling, all seven strategies were evaluated on the best-performing base configuration (LogNormal formulation with Gamma2 prior), varying only the rule applied when a pseudo-label violates the censoring constraint ($\hat{y}_{ij}$). The strategies are defined as follows:

- **Unconstrained VI:** Variance-inflation formulation (Eq. (15)). Constraint-violating pseudo-labels are retained as noisy observations; the inflated effective variance $\sigma^2 + s_{ij}^2$ absorbs the discrepancy without explicit correction.
- **Filtered VI:** Variance-inflation formulation. Constraint-violating pseudo-labels are removed entirely before likelihood construction.
- **Truncated Latent-Variable (RIPPLE):** Latent-variable formulation (Eq. (16)). A latent log-lifetime u_{ij} is introduced with truncation at $\ln \tau_i$, enforcing the censoring constraint probabilistically within the hierarchical model.
- **Violation-Censored:** When $\hat{y}_{ij} < \tau_i$, the pseudo-label is discarded and replaced by a standard right-censored survival likelihood at

Table 15

Cross-class predictive comparison between T-NT and R-TNT. Δ ELPD denotes the difference in expected log predictive density between R-TNT and T-NT. Avg indicates the average log predictive density across test samples.

Train $\rightarrow$ Test	T-NT	R-TNT	Δ ELPD	Avg T-NT	Avg R-TNT	Δ Avg
6.6 $\rightarrow$ 13.8	-1367.005	-79.065	1287.940	-136.701	-7.907	128.794
13.8 $\rightarrow$ 6.6	-145.497	-52.933	92.564	-20.785	-7.562	13.223

Table 16

Sensitivity analysis for different pseudo-label constraint handling rules. All configurations use the LogNormal formulation with Gamma2 prior; only the constraint handling rule is varied. VI = Variance Inflation (Eq. (15)); Truncated Latent-Variable = latent-variable formulation with truncation at τ_i (Eq. (16)).

Method	WAIC $\downarrow$	LOO $\downarrow$	MAPE $\downarrow$	MPE	R^2 $\uparrow$	max Pareto $\hat{k}$
Unconstrained VI	269.567	269.589	0.802	-0.708	0.896	0.264
Truncated Latent-Variable	268.412	268.505	0.453	-0.176	0.885	0.331
Filtered VI	269.677	269.698	0.798	-0.700	0.892	0.241
Soft-Penalized VI (0.5)	269.656	269.674	0.808	-0.714	0.892	0.186
Soft-Penalized VI (1)	269.817	269.835	0.816	-0.724	0.888	0.170
Soft-Penalized VI (2)	269.722	269.740	0.808	-0.715	0.889	0.271
Violation-Censored	272.555	272.574	1.041	-0.973	0.661	0.263

τ_i , retaining only the information that the specimen survived at least until τ_i .

- **Soft-Penalized VI (ρ):** Variance-inflation formulation with a violation-severity penalty. The log-likelihood contribution of each constraint-violating pseudo-label is multiplied by $v_{ij} = \exp(-\rho \cdot \max(\ln \tau_i - \ln \hat{y}_{ij}, 0))$, where ρ governs penalty sharpness. Non-violating pseudo-labels receive weight 1. Note that this violation-severity weight is conceptually distinct from the Weibull extension weight $1/(1 + s_{ij}^2/c)$, which penalizes based on prediction uncertainty rather than censoring-constraint violation severity.

As shown in Table 16, Truncated Latent-Variable achieves the best WAIC, LOO, and MAPE values across all strategies, confirming the advantage of probabilistic truncation for handling censoring-constraint violations. The four Variance-Inflation-based variants (Unconstrained VI, Filtered VI, and all three Soft-Penalized VI configurations) yield comparable performance, with WAIC and LOO clustered between 269.6 and 269.9, indicating that within the VI formulation class the specific treatment of violating pseudo-labels has limited impact on overall inference quality. Notably, the three Soft-Penalized VI variants show negligible sensitivity to ρ , and larger ρ does not yield monotonic improvement, suggesting that an exponential violation penalty offers no consistent advantage over either Unconstrained VI or Filtered VI. The Violation-Censored strategy incurs the highest WAIC, LOO, and MAPE (1.041), indicating that replacing pseudo-label information with a censored likelihood upon every violation leads to non-trivial information loss. Collectively, these results suggest that the critical factor is not how violations are handled within a Variance-Inflation framework, but rather whether the model adopts a principled hierarchical structure—as in Truncated Latent-Variable—that enforces the censoring constraint directly through the posterior.

4.4.3. Sensitivity to pseudo-label uncertainty

Fig. 5 shows the effect of scaling pseudo-label uncertainty on predictive behavior. The variance-inflation formulation exhibits strong sensitivity, leading to rapidly increasing credible intervals and unstable predictions. In contrast, the latent-variable formulation remains stable across a wide range of uncertainty scales. The increase in predictive uncertainty is gradual, and the posterior median is largely unaffected. This indicates that the model incorporates pseudo-label uncertainty in a controlled and well-calibrated manner.

4.4.4. Robustness to censoring ratio

Fig. 6 evaluates model robustness under increasing censoring ratios. The censored-likelihood baseline exhibits a dramatic increase in predictive uncertainty, reflecting the lack of failure information.

Table 17

Comparison of rated-voltage incorporation strategies using Bayesian information criteria and predictive metrics. Overall rank is the sum of metric-wise rank scores across four evaluation criteria with a lower value indicating consistently superior performance across all dimensions. The Additive (RIPPLE) formulation achieves the lowest Overall rank (= 4), ranking first across all evaluation criteria.

Formulation	WAIC $\downarrow$	LOO $\downarrow$	MAPE $\downarrow$	R^2 $\uparrow$	Overall rank $\downarrow$
Additive (RIPPLE)	268.412	268.505	0.453	0.885	4
Scaled log-additive	271.319	271.924	0.466	0.867	10
Shifted additive	268.915	269.034	0.477	0.839	14
Log-scaled additive	271.776	272.429	0.467	0.865	16
Scaled additive	271.635	272.133	0.473	0.854	16
Baseline ratio	272.540	273.179	0.504	0.744	24
Log-additive	292.237	292.358	1.056	0.162	28
Shifted log-additive	292.710	292.850	1.065	0.152	32

The variance-inflation formulation partially mitigates this issue but still shows unstable behavior, with large fluctuations in both uncertainty and median predictions. In contrast, the latent-variable formulation remains robust. Predictive uncertainty grows moderately, and median predictions remain stable even under high censoring levels. This demonstrates that the model effectively leverages pseudo-label information to compensate for missing failures.

4.4.5. Comparison of rated-voltage formulations

To examine how R should be incorporated into the lifetime model, an ablation study was conducted using the LogNormal formulation with the best-performing informative prior (Gamma2). Several alternative specifications of the stress-life relationship were evaluated, each expressed in terms of $\ln L$:

- Baseline ratio: $\ln L = \ln \theta_1 - \theta_2 \ln(V/R) + \theta_3/T$.
- Additive (RIPPLE): $\ln L = \ln \theta_1 + R - \theta_2 \ln V + \theta_3/T$.
- Log-additive: $\ln L = \ln \theta_1 + \ln R - \theta_2 \ln V + \theta_3/T$.
- Shifted additive: $\ln L = \ln \theta_1 + (\theta_4 + R) - \theta_2 \ln V + \theta_3/T$.
- Scaled additive: $\ln L = \ln \theta_1 + (\theta_4 R) - \theta_2 \ln V + \theta_3/T$.
- Shifted log-additive: $\ln L = \ln \theta_1 + (\theta_4 + \ln R) - \theta_2 \ln V + \theta_3/T$.
- Scaled log-additive: $\ln L = \ln \theta_1 + (\theta_4 \ln R) - \theta_2 \ln V + \theta_3/T$.
- Log-scaled additive: $\ln L = \ln \theta_1 + (\ln \theta_4)R - \theta_2 \ln V + \theta_3/T$.

For formulations that included the additional parameter term θ_4 , a weakly-informative prior $\Gamma(1.0, 0.01)$ was assigned.

Table 17 compares alternative formulations using Bayesian predictive criteria and point-prediction accuracy metrics. Overall rank is the sum of metric-wise rank scores across four evaluation criteria—WAIC,

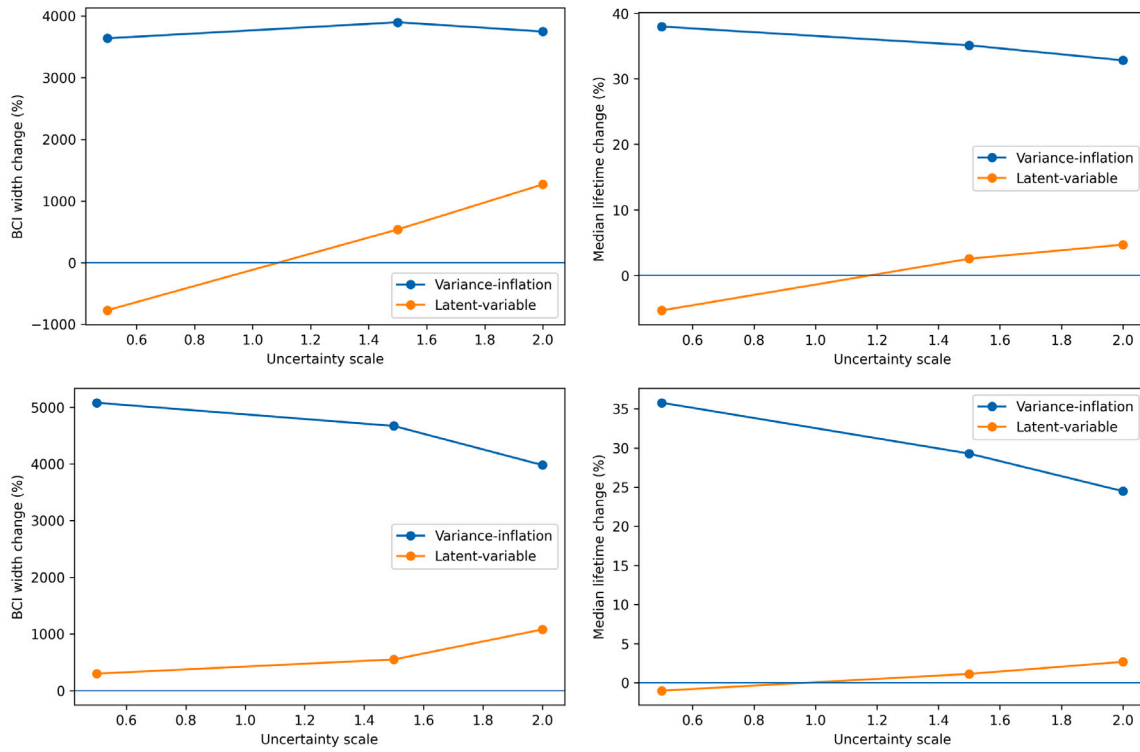


Fig. 5. Sensitivity of predictive uncertainty and median lifetime to pseudo-label uncertainty scaling. Top: 13.8 kV; bottom: 6.6 kV. Left: BCI width; right: posterior predictive median.

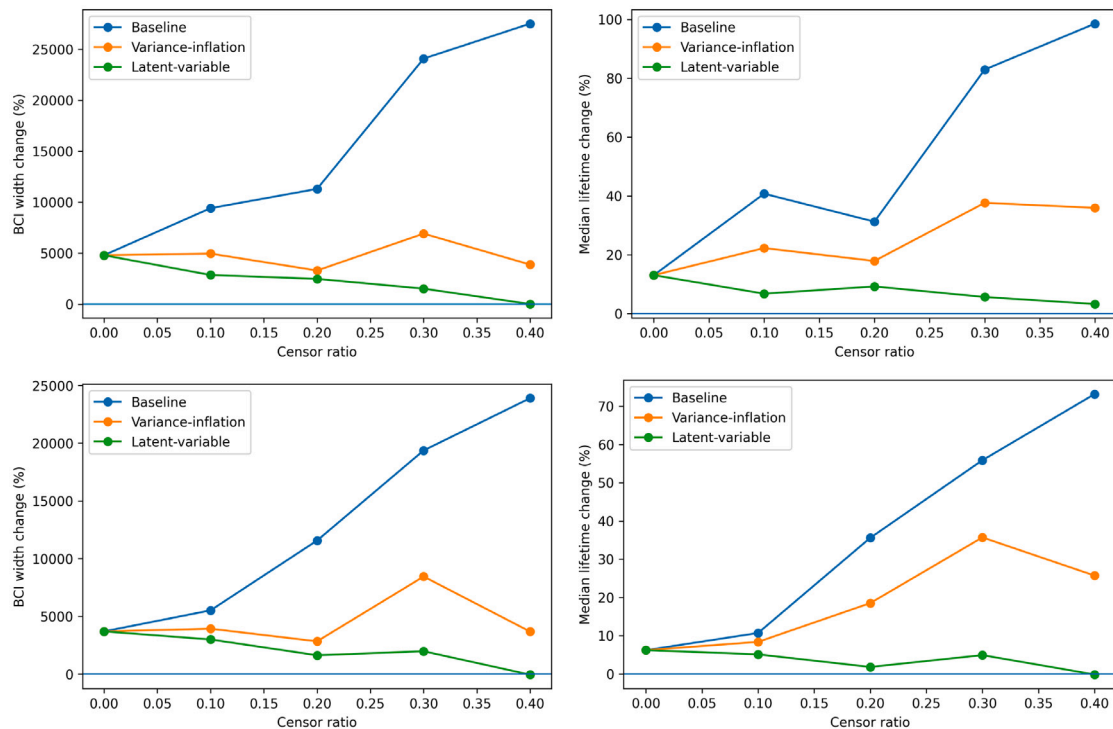


Fig. 6. Sensitivity of predictive uncertainty and median lifetime to censoring proportion. Top: 13.8 kV; bottom: 6.6 kV. Left: BCI width; right: posterior predictive median.

LOO, MAPE, and R^2 —with a lower value indicating consistently superior performance across all dimensions. The direct-additive formulation (RIPPLE) achieves the lowest overall rank, confirming its superiority across all criteria.

Logarithmic variants (Log-additive, Shifted log-additive) produce superficially competitive WAIC values but substantially degraded MAPE (1.056–1.065 vs. 0.453 for RIPPLE), because WAIC measures average log-predictive density rather than point-prediction accuracy: a biased

point estimate can yield acceptable WAIC if predictive variance is sufficiently wide. Physically, logarithmic compression of R distorts the discrete, additive nature of voltage-class design margins, driving systematic point-prediction errors. Scaled and shifted variants introduce unnecessary free parameters without recovering the underlying linear relationship, yielding inferior R^2 and intermediate overall ranks. These results confirm that the direct-additive entry of R into $\ln L$ is the most parsimonious and empirically effective representation under the fixed kV convention.

Scaled variants (Scaled additive, Scaled log-additive, Log-scaled additive) suffer from an analogous identification problem: the additional free parameter θ_4 partially absorbs the variance attributable to R , but cannot recover the underlying linear relationship between design class and log-lifetime, resulting in collectively inferior R^2 values and intermediate Overall ranks (10–32). The Baseline ratio formulation, which encodes voltage stress as $\ln(V/R)$, conflates the effect of operating voltage and rated-voltage into a single ratio and thereby loses the ability to disentangle insulation design margin from electrical loading, further degrading both MAPE and R^2 . Collectively, these results confirm that the direct additive entry of R into $\ln L$ is not merely a convenient simplification but a physically consistent encoding of the discrete, additive nature of voltage-class design margins.

4.5. Simulation

To further validate RIPPLE beyond the single real-world dataset, a comprehensive simulation study is conducted designed to reflect realistic ALT conditions.

4.5.1. Simulation setup

Three rated-voltage classes are considered, $R \in \{6.6, 11, 13.8\}$ kV, which are representative of medium-voltage industrial rotating machines. For each class, stress conditions are constructed using combinations of temperature (T) and voltage (V). For $R = 6.6$, stress levels are defined as (145 °C, 1.7 R), (155 °C, 1.9 R), and (165 °C, 2.1 R), while for $R = 11$ and 13.8, we use (145 °C, 1.7 R), (150 °C, 1.8 R), and (155 °C, 1.9 R). The simulation parameters are derived from posterior estimates obtained from the real dataset, ensuring that the generated data reflect realistic degradation behavior. Censoring proportions are varied across 20%, 30%, and 40%, and the number of samples per stress condition is varied from 5 to 10. For each configuration, 100 independent replications are performed.

Three data-generating mechanisms (DGMs) were evaluated: Scenario 1 (model-consistent) follows the assumed R–TNT additive structure; Scenario 2 (nonlinear R effect) replaces the additive R term with a logarithmic effect, introducing mild structural misspecification; Scenario 3 (interaction effect) adds a cross-class interaction between rated-voltage and relative electrical stress. Full DGM specifications and parameter settings are provided in Supplementary Material (S3).

4.5.2. Overall performance

Table 18 summarizes overall performance across all simulation settings. RIPPLE achieves the lowest MAPE (0.809) among all evaluated models, outperforming No pseudo-labels (1.316), No-R baseline (1.105), and Ratio baseline (1.359) by margins of 38.5%, 26.8%, and 40.5%, respectively. WAIC and LOO further corroborate this ranking, with RIPPLE recording substantially lower values than all baselines. The BCI width of RIPPLE (12,403.914) is approximately 3.7 times narrower than that of No pseudo-labels (45,721.554), confirming that pseudo-label integration substantially sharpens predictive intervals. While the No-R and Ratio baselines yield narrower BCI widths (6738 and 6746), their substantially higher MAPE values indicate that interval compactness comes at the cost of accuracy rather than reflecting genuine precision.

The observed coverage (0.813) falls below the nominal 0.90 level, reflecting the sharper predictive intervals induced by pseudo-label

Table 18

Overall simulation performance comparison across the evaluated models.

No pseudo-labels: R–TNT model with censored likelihood only (no pseudo-labels); serves as the no-pseudo-label baseline. *No-R baseline:* truncated likelihood with uncertainty-aware pseudo-labels; R covariate omitted from the stress-life model. *Ratio baseline:* truncated likelihood with uncertainty-aware pseudo-labels; voltage stress encoded as $\ln(V/R)$ rather than treating V and R as separate covariates.

Model	MAPE	Coverage ₉₀	BCI width	WAIC	LOO
RIPPLE	0.809	0.813	12403.914	715.475	715.565
No pseudo-labels	1.316	0.945	45721.554	733.464	733.517
No-R baseline	1.105	0.831	6738.322	743.544	743.607
Ratio baseline	1.359	0.827	6745.745	759.866	759.930

integration. This represents a deliberate trade-off between sharpness and calibration: while the no-pseudo-label baseline achieves higher coverage (0.945), its BCI width is approximately 3.7 times wider, limiting practical utility for engineering decision-making. RIPPLE is therefore presented as a method that improves predictive efficiency under severe censoring, with sub-nominal coverage acknowledged as a limitation.

4.5.3. Robustness to model misspecification

Model robustness under different DGMs is evaluated in Fig. 7. RIPPLE achieves its best performance in Scenario 1, where the data-generating mechanism exactly matches the assumed R–TNT structure. In this well-specified case, RIPPLE attains the lowest WAIC and MAPE among all models, confirming that it can accurately recover the underlying stress-life relationship when its structural assumptions are satisfied.

In Scenario 2, the rated-voltage effect follows a logarithmic rather than additive form, introducing systematic misspecification in the R–TNT formulation. Under this condition, the No-R baseline achieves a slightly better WAIC than RIPPLE, indicating that when the functional form of the rated-voltage effect is substantially misspecified, explicitly including R as a linear covariate may introduce bias that outweighs its benefit. In contrast, RIPPLE continues to outperform other baselines such as the ratio formulation and the no-pseudo-label model, and maintains competitive MAPE performance. This suggests that the pseudo-labeling mechanism and uncertainty-aware likelihood remain beneficial even when the structural role of R is imperfectly specified, while also highlighting the sensitivity of likelihood-based criteria to misspecified covariate effects.

In Scenario 3, which introduces interaction effects between rated-voltage and electrical stress, RIPPLE again achieves the best overall performance. This indicates that the model remains effective when the primary monotonic relationship between rated-voltage and lifetime is preserved, even if higher-order interactions are not explicitly modeled. In this case, the misspecification arises from omitted interaction terms rather than an incorrect functional form of the main effect, allowing RIPPLE to retain strong predictive performance. Across all scenarios, RIPPLE consistently outperforms baselines that either omit R entirely or encode it through simplified transformations such as $\ln(V/R)$, indicating that explicit modeling of rated-voltage as a design covariate provides substantial benefits when its effect is approximately well specified.

Overall, these results indicate that RIPPLE is robust to moderate model misspecification, particularly when the primary effect of rated-voltage is correctly specified. At the same time, the results reveal a predictable limitation: when the functional form of the rated-voltage effect itself is substantially misspecified (as in Scenario 2), explicitly modeling R may introduce systematic bias, and simpler models that omit the covariate can occasionally achieve better likelihood-based fit. This contrast highlights the difference between misspecification in the main effect and omission of higher-order interactions, and underscores

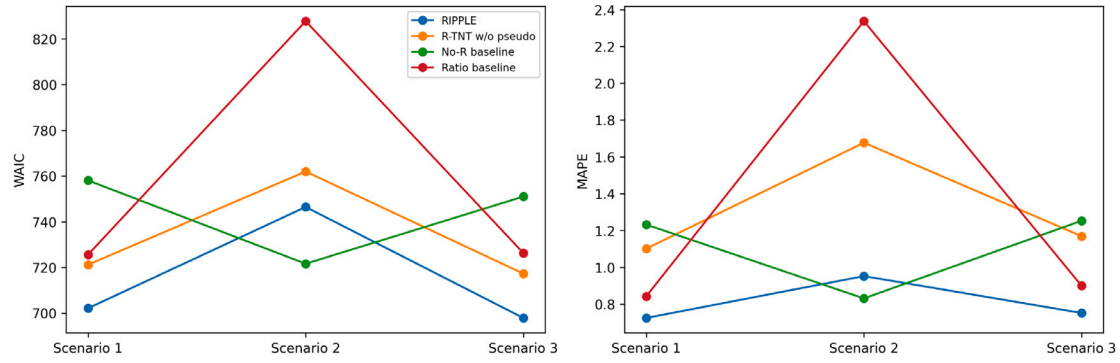


Fig. 7. Comparison of model performance (MAPE and WAIC) across different data-generating mechanisms. Lines represent the mean MAPE and WAIC of RIPPLE compared with baseline models under three synthetic scenarios.

Table 19

Cross-class generalization performance under interpolation and extrapolation settings.

Setting	Model	MAPE	Coverage ₉₀	BCI width	WAIC	LOO
Extrapolation	RIPPLE	0.727	0.771	18317.480	447.773	447.860
Interpolation	RIPPLE	0.419	0.780	1359.861	463.408	463.494
Extrapolation	No pseudo-labels	1.117	0.866	55917.026	461.274	461.316
Interpolation	No pseudo-labels	0.462	0.942	3220.781	476.634	476.677
Extrapolation	No-R baseline	1.629	0.593	12173.096	474.410	474.464
Interpolation	No-R baseline	0.575	0.911	2526.925	494.237	494.289
Extrapolation	Ratio baseline	2.130	0.521	4887.465	473.428	473.491
Interpolation	Ratio baseline	0.748	0.848	9775.961	496.702	496.765

the importance of appropriate functional specification of design variables. Nevertheless, the overall results support the practical utility of the R-TNT framework under realistic conditions where the true form is only approximately known.

4.5.4. Cross-class generalization

Table 19 evaluates generalization to unseen voltage classes under interpolation (test class: 11 kV, trained on 6.6 and 13.8 kV) and extrapolation (test class lies outside the training range) settings. Under interpolation, RIPPLE achieves the lowest MAPE (0.419) and WAIC (463.408), outperforming all baselines by 9.3–44.0% in MAPE, with substantially narrower BCI width (1359.861 vs. 3220.781 for No pseudo-labels). Under extrapolation, the performance gap widens further (RIPPLE MAPE: 0.727 vs. 1.117–2.130 for baselines), indicating that the structural benefit of explicit rated-voltage parameterization becomes increasingly important as the model is required to generalize beyond the training voltage classes. The No-R and Ratio baselines show markedly low coverage under extrapolation (0.593 and 0.521), confirming poor calibration in unseen voltage class settings. These findings collectively establish that the explicit incorporation of rated-voltage as a separate covariate is a key enabler of robust cross-class generalization.

4.5.5. Effect of pseudo-label integration

The impact of pseudo-label integration is examined through a stepwise comparison of model variants (Table 20). The stepwise comparison reveals a clear monotonic improvement in predictive accuracy as pseudo-label sophistication increases. No pseudo-labels, which relies solely on censored-likelihood information, achieves a MAPE of 1.316 with a BCI width of 45,721.554 and WAIC of 733.464. Introducing deterministic pseudo-labels (R-TNT, deterministic pseudo-labeling) reduces MAPE to 1.131—an improvement of 14.1% relative to the no-pseudo-label baseline—while also narrowing the BCI width to 24,858.371, a reduction of approximately 45.6%, and improving WAIC to 729.177. These gains confirm that even a naive pseudo-labeling strategy provides meaningful benefits over purely censored-likelihood estimation by supplying additional failure-time information for censored specimens.

The uncertainty-aware pseudo-labeling employed by RIPPLE achieves the most substantial gains across all metrics. MAPE is further reduced to 0.809, corresponding to an overall improvement of 38.5% over the no-pseudo-label baseline and 28.5% over deterministic pseudo-labeling. BCI width is compressed to 12,403.914—approximately 72.8% narrower than the no-pseudo-label baseline and 50.1% narrower than the deterministic pseudo-label variant. WAIC improves to 715.475, a gain of 18.0 units over the no-pseudo-label baseline and 13.7 units over deterministic pseudo-labeling, confirming that uncertainty-aware pseudo-labels enhance both the quality of in-sample fit and out-of-sample generalization capacity.

Coverage decreases from 0.945 (no pseudo-labels) to 0.935 (deterministic) to 0.813 (RIPPLE) as pseudo-label integration is progressively introduced. This reduction reflects the sharpening of predictive intervals described above and should be interpreted in conjunction with the large accompanying reductions in BCI width. The improved interval sharpness of RIPPLE reflects genuine distributional learning facilitated by informative pseudo-label signals, rather than calibration degradation, and the resulting intervals are substantially more actionable for engineering decision-making despite their modest departure from the nominal 0.90 coverage level. Taken together, these results demonstrate that uncertainty propagation in pseudo-label construction is not merely an incremental refinement but constitutes a critical driver of RIPPLE's overall predictive superiority. Additional experiments on simulation data are provided in Supplementary Material (S4).

5. Conclusion

Reliable lifetime prediction for insulation in high-voltage rotating machines remains challenging due to data scarcity, high censoring ratios, and the practical difficulty of conducting long-duration multi-stress tests. To address these challenges, this study introduced **RIPPLE** (Rated-voltage-Informed Pseudo-Probabilistic Lifetime Estimation), a Bayesian accelerated failure-time framework designed to improve lifetime inference and cross-class extrapolation under severe data scarcity.

The proposed framework integrates three core methodological components. First, a *deep learning-based pseudo-labeling* strategy was developed to generate failure-like lifetime targets for censored specimens

Table 20
Effect of pseudo-label integration and uncertainty-aware pseudo-labeling in simulation.

Model	MAPE	Coverage ₉₀	BCI width	WAIC	LOO
RIPPLE	0.809	0.813	12403.914	715.475	715.565
Deterministic pseudo-labeling	1.131	0.935	24858.371	729.177	729.216
No pseudo-labels	1.316	0.945	45721.554	733.464	733.517

using a lightweight Transformer-MLP model with ensemble-based uncertainty quantification, thereby increasing the effective information content of the likelihood. Second, the *R-TNT stress-life model* explicitly incorporated rated-voltage as a design covariate within a LogNormal AFT formulation, enabling joint inference and cross-class extrapolation across heterogeneous voltage classes. Third, a *truncated latent-variable likelihood* was introduced as the primary uncertainty-aware formulation to treat pseudo-labels as probabilistic soft evidence rather than exact observations, with a variance-inflation likelihood evaluated as an alternative. These components are integrated within a Bayesian inference framework that supports MCMC-based parameter estimation and principled uncertainty quantification under small-sample and heavily censored conditions.

Empirical validation on multi-stress ALT data from 6.6/13.8 kV form-wound coils demonstrates that RIPPLE substantially improves both predictive accuracy and uncertainty quantification compared to conventional censored-likelihood approaches, achieving markedly sharper credible intervals while maintaining reliable prediction performance. The framework further exhibits stable behavior across prior specifications and successfully generalizes across voltage classes. Simulation studies corroborate these findings, showing that RIPPLE consistently outperforms baseline models under varying sample sizes, censoring ratios, and data-generating mechanisms, with pseudo-label integration emerging as the primary driver of performance gains.

Several limitations should be acknowledged. First, the current dataset encompasses only two rated-voltage classes with a limited number of observed failures per stress condition, which restricts the statistical leverage for identifying between-class parameters and constrains the breadth of cross-class generalization claims. Second, although *R* serves as an effective class-level proxy for insulation design strength, it does not capture within-class variability arising from manufacturing tolerances or environmental influences. Third, pseudo-labeled lifetimes remain model-based statistical estimates and cannot fully substitute for direct failure evidence. Finally, the truncated latent-variable formulation produces sharper predictive intervals at the cost of moderate under-coverage (approximately 0.813 vs. the nominal 0.90 level), a trade-off that reflects the inherent tension between calibration and sharpness in probabilistic prediction.

Despite these limitations, the RIPPLE framework offers a practical and statistically grounded approach for data-efficient lifetime modeling of high-voltage insulation systems. By jointly leveraging specification-informed covariates, probabilistic pseudo-label augmentation, and Bayesian inference, RIPPLE extends reliability prediction beyond traditional single-class models and provides a flexible foundation for future investigations of multi-stress and multi-class reliability assessment in industrial electrical equipment.

CRedit authorship contribution statement

Sewon Kim: Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Donghwan Kim:** Writing – original draft, Methodology, Data curation, Conceptualization. **Joongki Oh:** Writing – original draft, Visualization, Data curation, Conceptualization. **Hyunsoo Yoon:** Writing – review & editing, Supervision, Project administration.

Declaration of generative AI use

The authors declare that no generative AI or AI-assisted technologies were used in the writing or editing of this manuscript, and that no images or figures were created using AI-based tools.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding: This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (RS-2025-02305884).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.res.2026.113005>.

Data availability

The data supporting the findings of this study are proprietary internal datasets and are not publicly available.

References

- [1] Kim H, Kong T, Lee SB, Kang T-J, Oh N, Kim Y, Park S, Lim C, Stone GC. Experience with stator insulation testing and turn/phase insulation failures in the power generation industry. *IEEE Trans Ind Appl* 2018;54(3):2225–36.
- [2] Ghassemi M. High power density technologies for large generators and motors for marine applications with focus on electrical insulation challenges. *High Volt* 2020;5(1):7–14.
- [3] Bilek V, Barta J, Toman M, Losak P, Bramerndorfer G. A comprehensive overview of high-speed solid-rotor induction machines: Applications, classification, and multi-physics modeling. *Int J Electr Power Energy Syst* 2025;166:110520.
- [4] Pascual F, Meeker W, Escobar L. Accelerated life test models and data analysis. In: Springer handbook of engineering statistics. Springer; 2006, p. 397–426.
- [5] Meeker WQ, Escobar LA, Pascual FG. Statistical methods for reliability data. John Wiley & Sons; 2021.
- [6] Ma Z, Li N, Chen X, Tan S, Zhang J, Fan Y, Song X, Huang Y. Impact of step voltage parameters on accuracy of evaluating XLPE insulation DC voltage endurance coefficient. *IET Gener Transm Distrib* 2023;17(19):4293–303.
- [7] Stone G, Gupta B, Lyles J, Sedding H. Experience with accelerated aging tests on stator bars and coils. In: IEEE international symposium on electrical insulation. IEEE; 1990, p. 356–60. <http://dx.doi.org/10.1109/ELINSL.1990.109772>.
- [8] Erto P, Giorgio M. Assessing high reliability via Bayesian approach and accelerated tests. *Reliab Eng Syst Saf* 2002;76(3):301–10. [http://dx.doi.org/10.1016/S0951-8320\(02\)00024-8](http://dx.doi.org/10.1016/S0951-8320(02)00024-8).
- [9] Liu B, Shi Y, Ng HKT, Shang X. Nonparametric Bayesian reliability analysis of masked data with dependent competing risks. *Reliab Eng Syst Saf* 2021;210:107502. <http://dx.doi.org/10.1016/j.res.2021.107502>.
- [10] Haq SU, Omranipour R. Accelerated life testing of high voltage stator coils with enhanced PET-mica insulation system. In: 2011 electrical insulation conference (EIC). IEEE; 2011, p. 479–82.
- [11] Kim H-C, Kim J-S. Development and validation of reliability testing methods for insulation systems in high-voltage rotating electrical machinery on ships. *J Mar Sci Eng* 2025;13(2):186.
- [12] Li C, Cheng J, Zhu H, Wen B, Zhao X, Kang W. Degradation modeling and remaining useful life prediction for electronic device under multiple stress influences. *Sci Rep* 2025;15(1):19117.
- [13] Paoletti G, Golubev A. Partial discharge theory and applications to electrical systems. In: Conference record of 1999 annual pulp and paper industry technical conference (cat. no. 99CH36338). IEEE; 1999, p. 124–38.
- [14] Li Y, Gao H, Chen H, Liu C, Yang Z, Zio E. Accelerated degradation testing for lifetime analysis considering random effects and the influence of stress and measurement errors. *Reliab Eng Syst Saf* 2024;247:110101. <http://dx.doi.org/10.1016/j.res.2024.110101>.

- [15] Zhang S, Zhai Q, Li Y. Degradation modeling and RUL prediction with Wiener process considering measurable and unobservable external impacts. *Reliab Eng Syst Saf* 2023;231:109021. <http://dx.doi.org/10.1016/j.ress.2022.109021>.
- [16] Fang Z, Yu X, Qin Y, Li D, Liu Q, Lu D, Wang K. Deterioration of microwave dielectric properties of low-loss thermosetting polyphenylene oxide/hydrocarbon resin induced by short-term thermo-oxidative aging. *Polym Degrad Stab* 2022;206:110193.
- [17] Sun P, Sima W, Jiang X, Zhang D, He J, Ye L. Review of accumulative failure of winding insulation subjected to repetitive impulse voltages. *High Volt* 2019;4(1):1–11.
- [18] IEC 60034-18-31:2021, functional evaluation of insulation systems—Test procedures for form-wound windings—Thermal evaluation and classification. 2021, Edition 1.0, International Electrotechnical Commission.
- [19] IEC 60034-18-32:2022, functional evaluation of insulation systems (type II)—Electrical endurance qualification procedures for form-wound windings. 2022, Edition 1.0, International Electrotechnical Commission.
- [20] IEC/TS 60034-18-33:2022, functional evaluation of insulation systems—Multifactor endurance under simultaneous thermal and electrical stresses. 2022, Technical Specification, Edition 1.0, International Electrotechnical Commission.
- [21] Kako Y, Kadotani K, Tsukui T. An analysis of combined stress degradation of rotating machine insulation. *IEEE Trans Electr Insul* 1983;EI-18(6):642–50. <http://dx.doi.org/10.1109/TEI.1983.298703>.
- [22] Montanari G, Cacciari M. A probabilistic insulation life model for combined thermal-electrical stresses. *IEEE Trans Electr Insul* 1985;EI-20(3):519–22. <http://dx.doi.org/10.1109/TEI.1985.348776>.
- [23] Elsayed EA, Zhang H. Design of PH-based accelerated life testing plans under multiple-stress-type. *Reliab Eng Syst Saf* 2007;92(3):286–92. <http://dx.doi.org/10.1016/j.ress.2006.04.016>.
- [24] McPherson JW. Reliability physics and engineering: Time-to-failure modeling. 2013.
- [25] Giangrande P, Madonna V, Nuzzo S, Galea M. Moving toward a reliability-oriented design approach of low-voltage electrical machines by including insulation thermal aging considerations. *IEEE Trans Transp Electrification* 2020;6(1):16–27.
- [26] Kulan MC, Baker NJ. Life-time characteristics of random wound compressed stator windings under thermal stress. *IET Electr Power Appl* 2019;13(9):1287–97.
- [27] Ramin R, Montanari GC, Yang Q. Designing the insulation system for motors in electrified aircraft: Optimization, partial discharge issues and use of advanced materials. *Materials* 2021;14(24):7555.
- [28] Bjurström A, Edin H, Hillborg H, Nilsson F, Olsson RT, Pierre M, Unge M, Hedenqvist MS. A review of polyolefin-insulation materials in high voltage transmission; from electronic structures to final products. *Adv Mater* 2024;36(52):2401464. <http://dx.doi.org/10.1002/adma.202401464>.
- [29] Pramualsingha S, Yutthagowith P. Engineering lifetime model of crosslinked polyethylene under electrical and thermal stresses. *Sensors Mater* 2021;33(7):2445–57.
- [30] Pang Z, Si X, Hu C, Du D, Pei H. A Bayesian inference for remaining useful life estimation by fusing accelerated degradation data and condition monitoring data. *Reliab Eng Syst Saf* 2021;208:107341. <http://dx.doi.org/10.1016/j.ress.2020.107341>.
- [31] Wang L, Pan R, Li X, Jiang T. A Bayesian reliability evaluation method with integrated accelerated degradation testing and field information. *Reliab Eng Syst Saf* 2013;112:38–47. <http://dx.doi.org/10.1016/j.ress.2012.11.010>.
- [32] Pan J, Sun B, Wu Z, Yi Z, Feng Q, Ren Y, Wang Z. Probabilistic remaining useful life prediction without lifetime labels: A Bayesian deep learning and stochastic process fusion method. *Reliab Eng Syst Saf* 2024;250:110313. <http://dx.doi.org/10.1016/j.ress.2024.110313>.
- [33] Chen C, Wang C, Guo J, Cui P, Zheng J, Liu Z. Remaining useful life prediction considering multiple uncertainty information via Bayesian BiGRU-based method. *Reliab Eng Syst Saf* 2025;264:111431. <http://dx.doi.org/10.1016/j.ress.2025.111431>.
- [34] Bell R, McWilliams D, O'donnell P, Singh C, Wells S. Report of large motor reliability survey of industrial and commercial installations. I. *IEEE Trans Ind Appl* 1985;21(4):853–64. <http://dx.doi.org/10.1109/TIA.1985.349532>.
- [35] Gjaerde A. A phenomenological aging model for combined thermal and electrical stress. *IEEE Trans Dielectr Electr Insul* 1997;4(6):674–80. <http://dx.doi.org/10.1109/94.654574>.
- [36] Montanari GC, Simoni L. Aging phenomenology and modeling. *IEEE Trans Electr Insul* 1993;28(5):755–76. <http://dx.doi.org/10.1109/14.237740>.
- [37] Cygan P, Laghari J. Models for insulation aging under electrical and thermal multistress. *IEEE Trans Electr Insul* 1990;25(5):923–34. <http://dx.doi.org/10.1109/14.59867>.
- [38] Bartnikas R, Morin R. Multi-stress aging of stator bars with electrical, thermal, and mechanical stresses as simultaneous acceleration factors. *IEEE Trans Energy Convers* 2004;19(4):702–14.
- [39] Ye X, Hu Y, Zheng B, Chen C, Zhai G. A new class of multi-stress acceleration models with interaction effects and its extension to accelerated degradation modelling. *Reliab Eng Syst Saf* 2022;228:108815. <http://dx.doi.org/10.1016/j.ress.2022.108815>.
- [40] Yi X, Wang Z, Liu S, Tang Q. Acceleration model considering multi-stress coupling effect and reliability modeling method based on nonlinear Wiener process. *Qual Reliab Eng Int* 2024;40(6):3055–78.
- [41] Dong Y, Zhou Z, Liu K, Yang X. Multi-stress accelerated degradation testing reliability assessment of LED lamp beads considering generalized coupling. *Appl Sci* 2024;14(19):8767.
- [42] Sun X, Cai W, Li M. A hierarchical modeling approach for degradation data with mixed-type covariates and latent heterogeneity. *Reliab Eng Syst Saf* 2021;216:107928. <http://dx.doi.org/10.1016/j.ress.2021.107928>.
- [43] Upadhyay S, Mukherjee B. Bayes analysis and comparison of accelerated Weibull and accelerated birnbaum-saunders models. *Commun Stat-Theory Methods* 2009;39(2):195–213.
- [44] Smit N, Raubenheimer L, Mazzuchi T, Soyer R. A Bayesian generalized erylring-Weibull accelerated life testing model. *Qual Reliability Eng Int* 2024;40(2):1110–25.
- [45] Li M, Meeker WQ. Application of Bayesian methods in reliability data analyses. *J Qual Technol* 2014;46(1):1–23.
- [46] Wang L, Pan R, Wang X, Fan W, Xuan J. A Bayesian reliability evaluation method with different types of data from multiple sources. *Reliab Eng Syst Saf* 2017;167:128–35. <http://dx.doi.org/10.1016/j.ress.2017.05.031>.
- [47] Ma Z, Wang S, Ruiz C, Zhang C, Liao H, Pohl E. Reliability estimation from two types of accelerated testing data considering measurement error. *Reliab Eng Syst Saf* 2020;193:106610. <http://dx.doi.org/10.1016/j.ress.2019.106610>.
- [48] Zhu R, Chen Y, Peng W, Ye Z. Bayesian deep learning for remaining useful life prediction: An active learning perspective. *Reliab Eng Syst Saf* 2022;228:108758. <http://dx.doi.org/10.1016/j.ress.2022.108758>.
- [49] Arias Chao M, Kulkarni C, Goebel K, Fink O. Fusing physics-based and deep learning models for prognostics. *Reliab Eng Syst Saf* 2022;217:107961. <http://dx.doi.org/10.1016/j.ress.2021.107961>.
- [50] ReliaSoft. Accelerated life testing: Reference. Tucson, AZ: ReliaSoft Publishing; 2024.
- [51] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: *Advances in neural information processing systems*. 30, 2017, URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [52] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. 2013, arXiv:1312.6034.
- [53] Dezfili H, Kelly D, Smith C, Vedros K, Galyean W. Bayesian inference for NASA probabilistic risk and reliability analysis. *Tech. rep.*, 2009.
- [54] Hespanhol L, Vallio CS, Costa LM, Saragiotto BT. Understanding and interpreting confidence and credible intervals around effect estimates. *Braz J Phys Ther* 2019;23(4):290–301.
- [55] Hoffman MD, Gelman A, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 2014;15(1):1593–623.
- [56] Abril-Pla O, Andreani V, Carroll C, Dong L, Fannesbeck CJ, Kochurov M, Kumar R, Lao J, Luhmann CC, Martin OA, et al. PyMC: a modern, and comprehensive probabilistic programming framework in python. *PeerJ Comput Sci* 2023;9:e1516.
- [57] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 2017;27(5):1413–32.