Results in Engineering 31 (2026) 111321

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

Probabilistic system-level prognostics in aerospace applications

Mariana Salinas-Camus, Oscar Carpentier, Nick Eleftheroglou *

Intelligent System Prognostics for Operations and Maintenance Group, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, Delft, 2629HS, the Netherlands

ARTICLE INFO

Keywords:
System-level prognostics
Remaining useful life
Uncertainty quantification
Aerospace structures

ABSTRACT

System-Level Prognostics (SLP) is essential for mission success, as it aims to predict the Remaining Useful Life (RUL) of an entire component rather than of its smaller structural subsystems, referred to here as coupons. Unlike coupon-level prognostics, SLP must capture degradation interactions among coupons while quantifying inherent uncertainties. Component-scale failure data are costly to obtain, and existing methods often rely on oversimplified assumptions or computationally intensive simulations. To address these challenges, this paper introduces the RUL Inoperabilities Model (RIM), a probabilistic framework inspired by the Inoperability Input-Output Model (IIM) that operates directly on coupon-level RUL predictions. The RIM is prognostic-model agnostic, interpretable, and data-efficient, requiring only coupon-level data to train base predictors and a single component-level degradation history for adaptation. By propagating probabilistic coupon predictions to the component level, RIM enables uncertainty-aware SLP. The method is validated on a three-coupon aluminum component using two different base predictors, Hidden Semi-Markov Model (HSMM) and physics-guided Particle Filter (PF), both trained only on single-coupon data, and consistently improves component-level RUL accuracy and uncertainty quantification over a naive baseline.

1. Introduction

Prognostics provide critical information for operational decision-making and are essential for ensuring the successful completion of assigned missions [1]. At the system level, this information is typically expressed in terms of the System Remaining Useful Life (SRUL), which supports decisions related to mission continuation, maintenance scheduling, and risk mitigation. Because predictions of future behavior are inherently uncertain, SRUL should be treated as a random variable rather than a deterministic quantity [2]. Inaccurate SRUL predictions can result in premature mission aborts or unexpected failures, both of which carry significant operational and safety consequences [3].

System-level prognostics (SLP) aims to predict the SRUL of a complete system by accounting for the interactions among its constituent subsystems, rather than treating each subsystem's Remaining Useful Life (RUL) in isolation. While this distinction is well recognized in the Prognostics and Health Management (PHM) community, the development of accurate and uncertainty-aware SLP methods remains a challenging and comparatively underexplored problem [3,4].

In aerospace applications, the system of interest is often the structure itself. Structural components connect all subsystems and frequently play a decisive role in determining whether a mission can continue safely. However, collecting structure-level degradation or failure data is rarely feasible due to cost, time, and safety constraints [5]. For instance, in civil aviation certification, only two full-scale structural tests are typically conducted: one quasi-static test and one fatigue test. As a result, structure-level prognostic models must often be developed under extreme data scarcity.

This limitation motivates approaches that can be trained using lower-scale experiments, such as coupon or element tests, while still producing reliable and uncertainty-aware predictions at the structural level. Bridging this scale gap is a central challenge for SLP in aerospace structures.

Existing prognostic approaches in PHM are commonly classified as model-based, data-driven, or hybrid methods [3]. Model-based approaches rely on physics-based degradation models and offer interpretability, but they require detailed system knowledge and parameter identification. Data-driven approaches are often more attractive for aerospace structures, where the complex behavior of composite materials and assemblies makes accurate physics-based modeling difficult. Hybrid approaches combine physics-based and data-driven elements to improve robustness and generalization.

Most existing PHM studies, however, focus on small-scale specimens such as composite coupons, which limits their direct applicability to full aerospace structures [5]. The building block approach addresses this issue by validating models at progressively increasing levels of structural complexity, from coupons to elements, subcomponents,

* Corresponding author.
E-mail address: n.eleftheroglou@tudelft.nl (N. Eleftheroglou).

https://doi.org/10.1016/j.rineng.2026.111321
Received 17 December 2025; Received in revised form 4 May 2026; Accepted 30 May 2026
Available online 4 June 2026

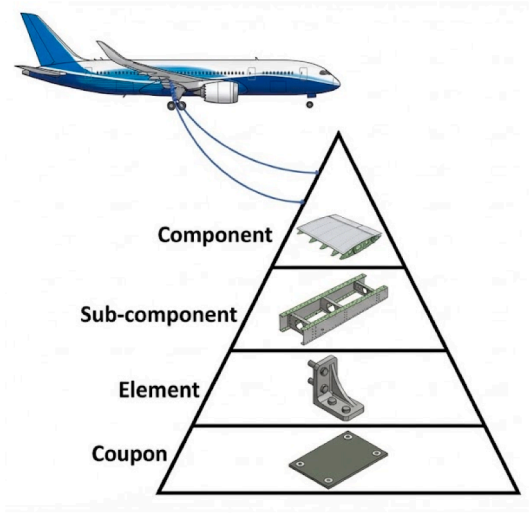


Fig. 1. Schematic representation of a PHM approach based on the building block approach inspired by [5].

and finally full structures, as illustrated in Fig. 1. For consistency with structural engineering terminology, the remainder of this paper uses the term coupon in place of subsystem and component in place of system.

Many existing SLP methods model a component either as a collection of independent coupons or as a black box driven solely by input-output data [3,4]. While computationally efficient, these approaches typically require large datasets and fail to capture interdependencies among coupons. Other methods assume that component failure occurs when the weakest coupon fails, which oversimplifies structural behavior, or impose predefined functional forms to represent coupon correlations, such as power-law relationships [6].

More comprehensive approaches explicitly model degradation interactions among subsystems, including physics-based formulations [7], ensemble learning methods [8], particle filtering techniques [7,9], agent-based models [10], Bayesian networks [11], and graph convolutional networks [12,13]. The Bayesian network-based approach proposed by Wang et al. [11], however, was primarily developed for risk assessment and fault diagnosis rather than for SRUL prognostics. Graph-based methods have also been proposed to capture subsystem interactions; for instance, Ifeanyi and Coble [13] employ simulated and normalized degradation data, resulting in a large number of degradation histories. Although compared against other deep learning techniques, the reported gains in predictive accuracy are limited, and uncertainty quantification is not addressed. Notably, two recent literature reviews on SLP [3,4] identify few studies proposing graph-based or Bayesian network models, indicating that such approaches remain underexplored in prognostics.

Overall, while these methods are powerful, they are often computationally demanding, data-intensive, and difficult to interpret, limiting their practical applicability in aerospace systems where component-level degradation data are scarce.

Nguyen et al. [14] note that SLP is itself rarely addressed in the literature and that explicit treatment of uncertainty within SLP is even less common. Although their proposed framework produces probabilistic SRUL estimates, it is demonstrated only for components composed of multiple independent coupons, which restricts its applicability to structural components where interdependencies among coupons are intrinsic. Similarly, Liu and Zio [15] obtain a probabilistic SRUL using a Bayesian framework combined with parallel Monte Carlo simulations. While their results are promising, a key assumption is that the dependency relationships between subsystems are known a priori and provided as input to the framework, which limits its applicability in practical structural prognostics.

An alternative framework that offers interpretability while explicitly representing interdependencies is the Inoperability Input-Output Model (IIM), originally developed in economics [16] and later adapted for engineering systems [17,18]. The IIM models interactions among system elements through a component matrix and requires relatively limited data. It is transparent, computationally efficient, and can be updated online as new information becomes available.

However, several limitations prevent the direct application of the IIM to aerospace SLP. First, it relies on inoperability measures that must be derived from physics-based degradation models, which are often unavailable or unreliable for complex aerospace structures. Second, it is not prognostic-model agnostic, as it assumes the existence of explicit degradation models rather than allowing arbitrary coupon-level predictors. Third, uncertainty propagation from coupon-level predictions to the SRUL is not addressed, despite its critical importance for decision-making.

To address these limitations, this paper introduces the RUL Inoperabilities Model (RIM), a novel probabilistic framework inspired by the principles of the IIM that operates directly on coupon-level RUL predictions. The RIM is prognostic-model agnostic and eliminates the need to predefine coupon interaction mechanisms. By propagating the probabilistic outputs of coupon-level predictors to the component level, the RIM enables uncertainty-aware SRUL estimation using only coupon-level training data and limited component-level degradation history. The proposed methodology is validated through a realistic case study involving non-linear coupon interdependencies, demonstrating its potential as a scalable, interpretable, and data-efficient solution for system-level prognostics of aerospace structures.

The contributions of this paper are as follows:

1. A novel and interpretable system-level prognostics framework (RIM) that models interdependencies among structural coupons while operating directly on coupon-level RUL predictions.
2. A probabilistic and prognostic-model-agnostic formulation that explicitly propagates uncertainty from the coupon level to obtain uncertainty-aware SRUL estimates.
3. Experimental validation on a structural case study demonstrating the scalability and suitability of the proposed approach for aerospace structures where component-level degradation data are scarce.

The remainder of this paper is organized as follows. Section 2 presents the methodology for implementing the RIM, including both offline training and online updating, and describes how uncertainty is propagated from the coupon level to the system level. Section 3 introduces the case study used to validate the proposed approach. The in-house experimental campaign is detailed, and coupon-level RUL predictions obtained using two distinct base prognostic models are presented to demonstrate the prognostic-model-agnostic nature of the RIM. The resulting system-level predictions are then evaluated in terms of both accuracy and uncertainty quantification. Finally, conclusions are presented in Section 4.

2. Methodology

The RUL Inoperabilities Model (RIM) proposed in this paper is designed to be broadly applicable to any engineering system that can be represented as a collection of subsystems. Since this study focuses on structures, the term coupon is used to denote a subsystem, while the term component refers to the overall system.

RIM extends on the Inoperability Input-Output Model (IIM) through several key adaptations. Most importantly, RIM is prognostic model-agnostic, as it directly uses coupon-level RUL predictions as inputs. The prognostic model applied at the coupon level is referred to as the base predictor. To demonstrate that RIM performs consistently regardless of the prognostic approach, two different base predictors are employed

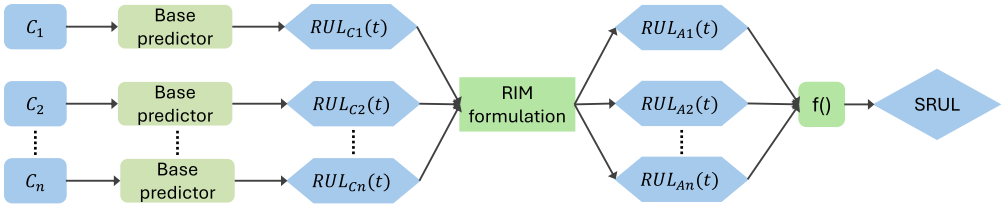


Fig. 2. General framework of the proposed RIM for SRUL predictions.

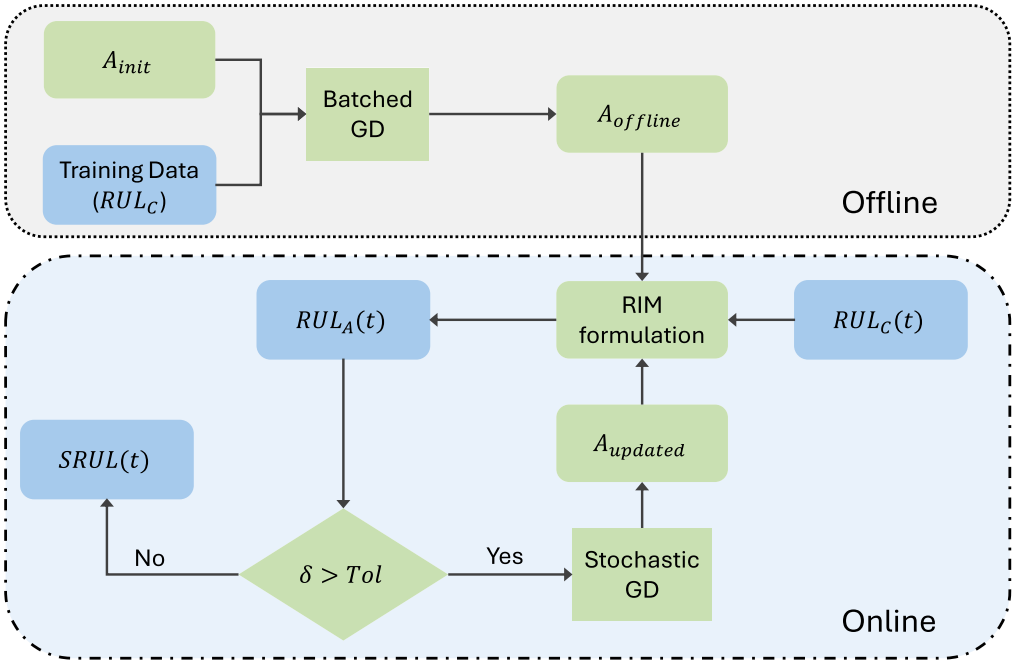


Fig. 3. Flowchart of the adapted RIMs online and offline workings.

in this study. The first is a Hidden Semi-Markov Model (HSMM) [19], which represents a data-driven method and is described in Section 3.2.1. The second is a physics-guided Particle Filter (PF), which exemplifies a model-based approach and is detailed in Section 3.2.2. Both base predictors are trained on data at the coupon level (not at the scale of interest). During testing, the models developed at the coupon scale are used to estimate the RUL at the component scale (the scale of interest).

The RIM framework, together with its training methodology, is presented in Section 2.1. An overview of the complete SLP framework, from coupon-level data to system-level RUL (SRUL) prediction, is shown in Fig. 2. In the Figure, $C_1, \ldots, C_n$, denote the condition monitoring (CM) data collected from the component coupons. The base predictors, previously trained on coupon data, are applied to these measurements during operation to produce the coupon-level RUL predictions, expressed as $\text{RUL}_{C1}, \ldots, \text{RUL}_{Cn}$. At each timestep t, these predictions are transformed by the RIM into adapted RULs, denoted as $\text{RUL}_{A1}, \ldots, \text{RUL}_{An}$, which account for interdependencies among the coupons. Finally, the SRUL can then be derived from the adapted RULs in a manner consistent with the components behavior. Depending on the application, the operator $f(\cdot)$ could correspond to the minimum, maximum, mean, or any other operator deemed appropriate.

2.1. RUL inoperabilities model (RIM)

In this work, the RIM model is developed so it can be used with any type of prognostic model as a base predictor and quantify RUL uncertainty at the component level. The RIM operates in two stages: an *offline* training phase and an *online* testing phase. A schematic overview of the process is provided in Fig. 3.

In the offline (training) phase of the RIM, base prognostic predictors estimate the RUL for each coupon, denoted collectively as $\text{RUL}_C(t)$, which represents the set $\text{RUL}_{C1}(t), \text{RUL}_{C2}(t), \ldots, \text{RUL}_{Cn}(t)$. These predictions, along with an initial estimate of the interdependency matrix A_{init}, are used to construct the initial RIM model. The matrix A_{offline} is learned using a training dataset composed of RUL predictions generated by the base predictors across the coupons of the full component.

It is important to note that the base predictors are trained on coupon data and are not exposed to component-level data during their training. As a result, the dataset used to estimate A_{offline} reflects how independently trained coupon-level models behave when applied to full-component data.

In the online (testing) phase, the model operates in real time, continuously receiving updated $\text{RUL}_C(t)$ predictions for each coupon. These are transformed into adjusted predictions, denoted collectively as $\text{RUL}_A(t)$, via the interdependency matrix A, where $\text{RUL}_A(t)$ represents the set of adjusted RULs across all coupons.

A key feature of the online phase is adaptive training: if the deviation $\delta = |\text{RUL}_A(t) - \text{RUL}_A(0)|$ exceeds a threshold *Tol*, the matrix A is updated to reflect the component's evolving behavior. The final SRUL is then computed by using an aggregation operator, $f(\cdot)$, on the vector $\text{RUL}_A(t)$.

2.1.1. Mathematical formulation

The original IIM formulation from [17] is given by:

$$q(t) = \kappa(t) \cdot [A \cdot q(t-1) + c(t)], \tag{1}$$

where $q(t)$ represents the vector of component-wide inoperabilities, $\kappa(t)$ captures environmental effects, and $c(t)$ denotes internal

inoperabilities. For application to aerospace structures, two modifications are made:

1. The environmental factor $\kappa(t)$ is excluded due to the complexity of modeling dynamic flight environments and the lack of representative environmental datasets.
2. Instead of inoperabilities, the model operates directly on the predicted RULs.

A key aspect of component degradation is that the condition of one coupon often affects the degradation of others. This interdependency motivates the creation of the relative degradation term, $z(t)$, which allows the model to account for interactions between coupons while leaving the explicit quantification of these interactions to the A-matrix.

Therefore, the proposed RIM is defined as:

$$\mathrm{RUL}_A(t) = \mathrm{RUL}_C(t) \circ (A \cdot z(t)) + \mathrm{RUL}_C(t), \tag{2}$$

where $\circ$ denotes the Hadamard (element-wise) product. The adapted RUL, $\mathrm{RUL}_A(t)$, modifies the coupon RUL, $\mathrm{RUL}_C(t)$, to account for interdependencies among coupons encoded in the matrix A. The term $\mathrm{RUL}_C(t) \circ (A \cdot z(t))$ provides a correction that is added to the baseline RUL, where $A \cdot z(t)$ acts as a multiplicative factor modulating each coupon's adjustment based on the relative states of all coupons. The vector $z(t)$ encodes the relative degradation of the coupons:

$$z(t) = \frac{\mathrm{RUL}_C(t)}{\frac{1}{N}\sum_{i=1}^{N} \mathrm{RUL}_{C(t)}^{(i)}}. \tag{3}$$

Each element of $z(t)$ reflects how an individual coupon's degradation compares to the component average. Values greater than one indicate better-than-average condition, while values below one indicate accelerated degradation. Through the matrix multiplication $A \cdot z(t)$, the relative states of all coupons influence each coupon's adjusted RUL according to their interdependencies.

2.1.2. Gradient descent for matrix training

The matrix A is the only trainable parameter in the model and is learned using gradient descent (GD). Two GD variants are used:

- **Batch GD** during offline training for stable convergence.
- **Stochastic GD** during online training for faster and more responsive updates.

The model is trained using the Mean Squared Error (MSE) loss:

$$\mathcal{L}(\mathrm{RUL}_A(t), \mathrm{RUL}_{\mathrm{true}}) = \frac{1}{N}\sum_{i=1}^{N}\left(\mathrm{RUL}_A^{(i)}(t) - \mathrm{RUL}_{\mathrm{true}}^{(i)}\right)^2. \tag{4}$$

The gradient of the loss for each matrix entry a_{ij} is:

$$\frac{\partial \mathcal{L}}{\partial a_{ij}} = \frac{2}{N} \cdot \mathrm{RUL}_C^{(i)} \cdot \left(\mathrm{RUL}_A^{(i)} - \mathrm{RUL}_{\mathrm{true}}^{(i)}\right) \cdot z^{(j)}. \tag{5}$$

The matrix update rule is:

$$a_{ij}^{\mathrm{new}} = a_{ij}^{\mathrm{prev}} - \lambda \cdot \frac{\partial \mathcal{L}}{\partial a_{ij}}, \tag{6}$$

where λ is the learning rate.

2.1.3. Offline phase

In the offline phase, true RUL values are available as labels. The matrix A_{init} is zero-initialized to reflect the absence of prior knowledge regarding interactions between coupons. The matrix is then updated iteratively using batch GD. Owing to noise and gradient averaging effects inherent to batch optimization, convergence-based stopping criteria are unreliable. Consequently, a fixed number of training iterations is adopted, empirically selected via cross-validation on the training datasets.

2.1.4. Online phase

Since true RUL values are unavailable during online training, an alternative labeling method is required. This approach compares past predictions to the current $\mathrm{RUL}_C(t)$ using a timestep difference k, enabling the model to learn changes over extended intervals. To minimize computational overhead during operation, continuous retraining is avoided. Instead, training is triggered only when the prediction error at the last update exceeds a defined threshold. Once triggered, the A matrix is updated via stochastic GD for a fixed number of iterations.

After computing the adapted RULs of each coupon, denoted as $RUL_A(t) = [RUL_{A1}(t), \ldots, RUL_{An}(t)]$, the SRUL can be obtained by aggregating these adapted RULs. Formally, the SRUL at time t is defined as:

$$SRUL(t) = f(RUL_A(t)), \tag{7}$$

where $f(\cdot)$ represents an aggregation operator chosen to reflect the behavior of the component. Depending on the configuration and safety requirements, $f(\cdot)$ could correspond to the minimum, maximum, mean, or another statistic that best captures the components expected failure behavior. This formulation provides flexibility, allowing the SRUL calculation to be tailored to different applications.

2.1.5. Uncertainty quantification in RIM

Uncertainty quantification is essential in prognostics, since point estimates of RUL are insufficient for reliable decision-making [2]. In the RIM framework, uncertainty present at the coupon level is propagated to the SRUL through a sampling-based approach.

At each time step, M samples are drawn from the probability distribution functions (PDFs) provided by the base prognostic models for each coupon. These PDFs represent the uncertainty captured by the base predictors. In principle, they may include both aleatoric and epistemic uncertainty, though in practice, base predictors often capture only part of this uncertainty. A stratified sampling strategy [20] is used to ensure adequate coverage of the distributions while controlling computational cost.

Each sample consists of one RUL value per coupon, forming a complete input tuple:

$$\mathrm{Sample}_m = \left(\mathrm{RUL}_{C1}^{(m)}(t), \mathrm{RUL}_{C2}^{(m)}(t), \ldots, \mathrm{RUL}_{Cn}^{(m)}(t)\right), \tag{8}$$

where $m = 1, 2, \ldots, M$ and n is the number of components.

For each sample, Eq. 2 is applied to compute the corresponding adjusted RUL vector $\mathrm{RUL}_A^{(m)}(t)$. From this, the SRUL is derived by applying an aggregation operator (analogous to Eq. 7):

$$\mathrm{SRUL}^{(m)}(t) = f\left(\mathrm{RUL}_A^{(m)}(t)\right). \tag{9}$$

Repeating this procedure across all M samples yields an empirical probability distribution of the SRUL at time t. This distribution captures the propagated uncertainty from the coupon-level models, providing both a mean estimate and confidence intervals for the SRUL.

It should be noted that this procedure propagates only the uncertainties encoded in the coupon-level RUL distributions. Uncertainties related to the learned interdependency matrix A or to the structural assumptions of the RIM itself are not explicitly represented.

3. Case study

The RIM is evaluated using a conceptual aerospace component composed of three interconnected open-hole rectangular coupons. Model training begins with single-coupon fatigue data, then extends to the multi-coupon component to assess the RIMs ability to predict the RUL of higher-level components from lower-level coupons. These components were subjected to fatigue testing under conditions similar to those used in single-coupon tests, and the same type of CM data was collected during operation.

(a) Single open-hole coupon. (b) Component.

Fig. 4. Coupon and component composed of 3 interconnected open-hole coupons.

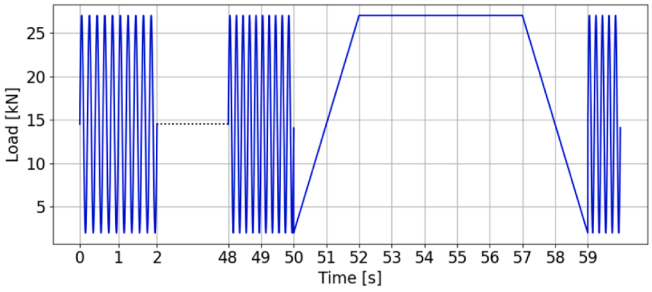


Fig. 5. Loading diagram of the programmed cycles of the fatigue machine.

3.1. Experimental campaign

To apply the proposed RIM approach in a realistic setting, a suitable experiment was chosen. Fatigue data from aluminum single open-hole coupons (Fig. 4a) formed the training set, while testing was performed on components composed of 3 single interconnected open-hole coupons (Fig. 4b).

All components were fabricated from 7075-T6 aluminum with a thickness of 2 mm. To reduce fatigue life, small notches were introduced around the holes, following the same approach as in the single-coupon configuration. Fatigue tests were conducted using an MTS machine with a 60 kN maximum load capacity.

CM data was obtained using Digital Image Correlation (DIC) to track the surface strain field during operation. In structural applications, strain serves as a reliable health indicator, as it directly correlates with stiffness degradation.

Single open-hole coupons were tested under constant-amplitude fatigue loading with a load ratio of 0.1, an amplitude of 9 kN, and a frequency of 5 Hz. In the case of the component tests, since each component consisted of three such coupons, the load amplitude was increased to 27 kN while maintaining the same load ratio and frequency. After 250 fatigue cycles, a ramp-up to the maximum load was applied and held for five seconds to allow image capture, which was used for strain calculations from the DIC. The loading process is shown in Fig. 5.

Axial strain data obtained from DIC were used for CM (see Fig. 6). The strain values were averaged per coupon for each frame, and truncated to include only the degradation phase (from crack initiation to failure). To prevent overestimation, strain values above the materials yield strain (0.007) were clipped. A zero-offset correction was applied by subtracting the initial strain value, providing a consistent undegraded baseline across all samples.

The training process was performed at a coupon scale rather than at a component level, using strain data from four open-hole rectangular coupons and a single component test. Fig. 7 illustrates the processed strain histories used for training the base predictions and validation of the RIM. In subfigure 7a, the data from the single open-hole specimens is presented, which is used to train the base predictors. Subfigure 7b shows the train component strain evolution, which served as the basis for training the RIM model. The remaining subfigures (7c–7f) display the test components. These test components were used to evaluate the system-level prognostic performance.

3.2. Base predictors

By using the data from single open-hole coupons, the base predictors were trained. As mentioned earlier, two predictors are considered to demonstrate how the RIM framework can generate SRUL predictions independently of the base prognostic model. A data-driven approach is represented by the Hidden Semi-Markov Model (HSMM), while a model-based approach is implemented through a Particle Filter with an exponential degradation model. Both base predictors were trained on degradation histories obtained from four open-hole coupons.

3.2.1. Hidden semi-Markov model (HSMM)

The Hidden Semi-Markov Model (HSMM) [19] extends the classical Hidden Markov Model (HMM) [21] by explicitly modeling the duration that an asset remains in each hidden state. This feature makes the HSMM particularly suitable for degradation processes, where transitions between degradation states progress gradually rather than at every observation step.

In this framework, the HSMM jointly describes (i) the probabilistic transitions between hidden degradation states, (ii) the distribution of observations associated with each state, and (iii) the distribution of state durations. Together, these parameters allow the model to capture realistic degradation trajectories and to infer the underlying state sequence from observed data. Once trained, the model can be used for three tasks: evaluating sequence likelihoods, decoding the most probable state sequence, and estimating model parameters from degradation histories.

For prognostics, three simplifying assumptions are applied. First, the asset always begins in the “healthy” initial state. Second, degradation

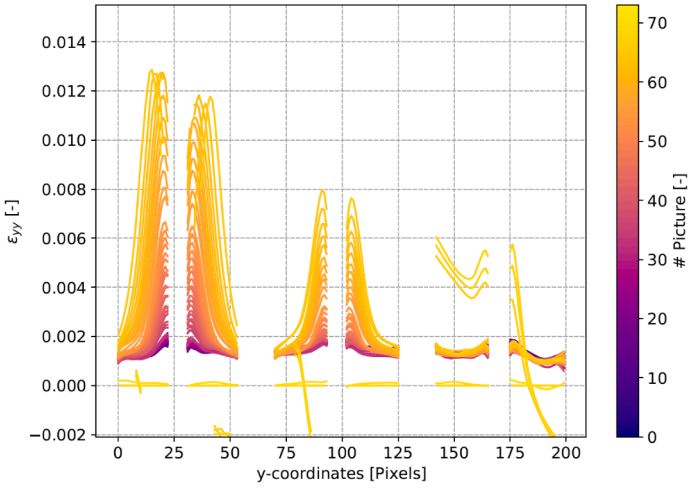
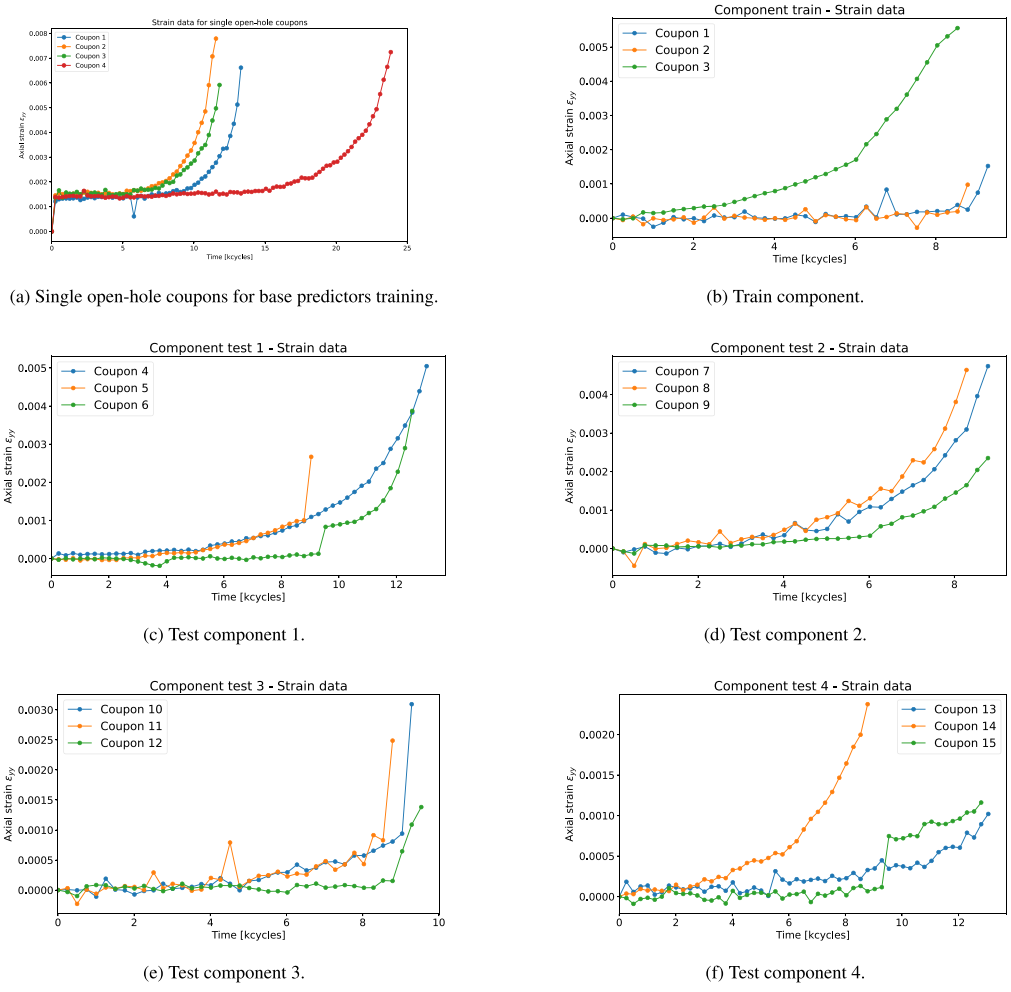


Fig. 6. Extracted strain data from DIC measurement.



(a) Single open-hole coupons for base predictors training.

(b) Train component.

(c) Test component 1.

(d) Test component 2.

(e) Test component 3.

(f) Test component 4.

Fig. 7. Strain data for single open-hole coupons, train component, and five test components.

progresses monotonically through left-to-right transitions into neighboring states. Third, the final failure state is absorbing and directly observable. Under these assumptions, the Viterbi algorithm can track degradation in real time, and RUL is predicted based on the inferred health state and its expected duration.

The HSMM is chosen as the data-driven base predictor because of the demonstrated success of HMM-based approaches in the literature, where stochastic models have shown strong capability in predicting the RUL of engineering systems by effectively modeling degradation dynamics [2,22–27]. Specifically, the HSMM performs reliably with limited training data and, due to its probabilistic nature, quantifies aleatoric uncertainty in RUL predictions. These qualities make the HSMM a strong and well-motivated representative of data-driven prognostics within the RIM framework.

3.2.2. Physics-guided particle filter

The physics-guided Particle Filter (PF) combines a parametric exponential degradation model with Bayesian filtering to predict RUL. An exponential model is used because strain-cycle data in aluminum alloys have been shown to follow an exponential trend [28].

The strain evolution of each coupon is described by a three-parameter exponential function:

$$\epsilon(N) = \epsilon_0 \cdot e^{\beta N} + \epsilon_\infty, \tag{10}$$

where N is the number of load cycles, ϵ_0 is the initial strain amplitude, β is the degradation rate, and ϵ_∞ is the asymptotic strain. Parameters $(\epsilon_0, \beta, \epsilon_\infty)$ are estimated from training data using nonlinear least-squares fitting. The mean and standard deviation of these estimates define log-normal priors for PF initialization, ensuring all parameters remain physically meaningful.

The PF is initialized with 5000 particles, each representing a candidate parameter set with equal weight. At each observation cycle, the predicted strain of each particle is compared with the measured strain, and particle weights are updated assuming Gaussian measurement noise ($\sigma = 5 \times 10^{-4}$). Systematic resampling is applied whenever particle weights become too uneven.

For RUL prediction, each particle is propagated forward until the predicted strain reaches the failure threshold. The difference between this predicted failure cycle and the current cycle gives a particle-specific RUL. The collection of RULs forms a posterior distribution, from which the mean and confidence intervals (5th and 95th percentiles) provide point estimates and uncertainty bounds.

The PF is used as the model-based base predictor because it combines a physically meaningful degradation model with Bayesian filtering, which helps capture and manage uncertainty. In addition, PFs have been applied in physics-guided prognostic frameworks using different degradation models across a range of applications [29,30].

3.3. RIM results

Having been established with two base RUL predictors, the RIMs ability to adapt coupon-level RUL predictions to SRUL predictions is evaluated. Two RIM configurations are considered: one with the HSMM (3.3.1) and another with the physics-guided PF (3.3.2). Both predictors were trained on four single open-hole coupons.

For offline training, data from one component were used, while the remaining four components were reserved for testing. This setup reflects realistic conditions in large aerospace structures, where typically only one component at the relevant scale can be tested to failure due to cost and experimental constraints.

Within the RIM framework, the operator $f(\cdot)$ is chosen as $\max()$ because the case study involves a parallel component configuration. This corresponds to the assumption that the system fails when the last coupon reaches failure.

To provide a baseline for comparison, a method called "Naive System RUL" (Naive SRUL) was implemented. This method computes the system RUL as the maximum of the coupon-level RULs at each time step, representing a simple but intuitive approach for parallel structures. For each RIM configuration (HSMM-based and PF-based), performance is compared with the Naive SRUL in terms of RMSE (mean prediction error), CRPS [31], coverage [2], and Normalized Mean Prediction Interval Width (NMPIW) [32].

The CRPS jointly accounts for prediction accuracy and uncertainty and is therefore regarded as a more comprehensive metric for assessing probabilistic models [33]. The coverage metric quantifies the proportion of true system RUL values that fall within the predicted confidence bounds, reflecting the reliability of the uncertainty estimates. Finally, the NMPIW measures the normalized width of the prediction intervals, enabling a fair comparison of uncertainty sharpness across different models.

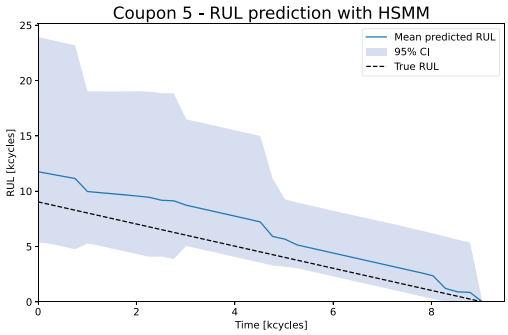


Fig. 8. Coupon level prediction using HSMM for test coupon 5.

Table 1
Hyperparameters for the offline and online phase of the RIM using HSMM as base predictor.

Type	Learning rate	Iterations	Batch	k
Offline	1e-7	2000	True	-
Online	1e-5	100	True	3

3.3.1. RIM with HSMM as base predictor

The HSMM was trained using degradation histories obtained from four single open-hole coupons (Subfigure 7a). The model was implemented in Python using the HiMAP library [34]. The number of hidden states was determined by evaluating candidate models using the Bayesian Information Criterion (BIC) [35], which selected four hidden states as optimal.

The trained HSMM was then used to predict RUL at the coupon level within each component. Taking the ground-truth as the coupon RUL, the average RMSE across 15 coupons (from 5 components) was 7.5 kcycles. Overall, the HSMM provided smooth and stable RUL trajectories with relatively low error, suggesting that it is a robust base predictor for the RIM. An example of RUL prediction at the coupon level is shown in Fig. 8.

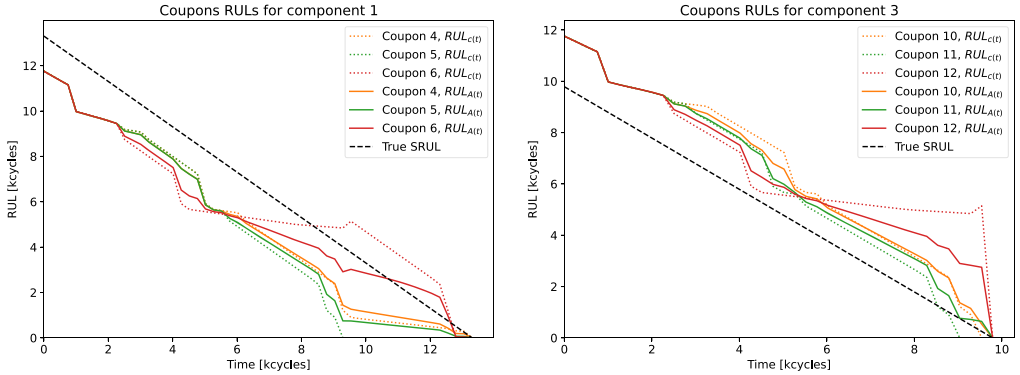
The selected hyperparameters for both offline and online phases are summarised in Table 1. The parameter k was set to 3, as larger values reduce accuracy by making past predictions less representative of the current degradation state. The learning rate was fixed at 1e-5 to ensure stable gradient descent and reliable convergence.

Fig. 9 presents coupon-level RUL predictions from the HSMM (dotted lines) alongside the adapted RUL predictions after applying the RIM formulation (solid lines) for component 1 (Fig. 9a) and component 3 (Fig. 9b). Notably, for coupon 12 (red), the RIM corrects the late-stage RUL prediction, mitigating overestimation. The results for both the Naive SRUL and the RIM are shown in Table 2.

The comparison between the RIM and the Naive SRUL approach across four test components reveals consistent benefits. RMSE values improved for all components, with a mean reduction from 7.58 to 6.82 kcycles.

The CRPS metric, which reflects both accuracy and confidence in probabilistic predictions, provides further evidence of the RIMs advantage. The mean CRPS decreased from 6.75 to 4.52, confirming that the RIM not only improves prediction accuracy but also generates better-calibrated uncertainty bounds.

Table 3 reports the empirical coverage and NMPIW for both the baseline and the RIM. Overall, the RIM preserves or slightly improves coverage while consistently reducing the interval width. The mean coverage increases marginally from 0.92 to 0.93, indicating that the adaptive correction does not compromise the reliability of the prediction intervals. In particular, for component 3, the RIM improves coverage from 0.82 to 0.88, suggesting better alignment between the predicted uncertainty and the observed degradation behavior.



(a) Component 1. (b) Component 3.

Fig. 9. Predicted coupon mean RULs and adapted mean RULs with RIM for each coupon for two components using HSMM as base predictor.

Table 2
RMSE and CRPS for baseline (Naive SRUL) and RIM across four components using HSMM as base predictor.

Component	RMSE [kcycles]		CRPS	
	Baseline	RIM	Baseline	RIM
1	5.56	**5.45**	4.97	**4.94**
2	8.69	**7.87**	8.19	**3.39**
3	9.94	**7.84**	8.18	**4.76**
4	6.15	**6.13**	5.65	**4.98**
Mean	7.58	**6.82**	6.75	**4.52**

Table 3
Coverage and NMPIW for baseline (Naive SRUL) and RIM across four components using HSMM as base predictor.

Component	Coverage		NMPIW	
	Baseline	RIM	Baseline	RIM
1	1.00	1.00	0.46	0.42
2	1.00	1.00	0.77	0.74
3	0.82	0.88	0.70	0.66
4	0.85	0.85	0.46	0.42
Mean	0.92	0.93	0.60	0.56

At the same time, the NMPIW decreases for all components, with the mean value reducing from 0.60 to 0.56. This indicates that the RIM yields systematically tighter prediction intervals without sacrificing coverage. For components 1 and 4, the reduction in NMPIW is achieved while maintaining full or near-constant coverage, highlighting the improved sharpness of the RIM SRUL predictions.

Fig. 10 further illustrates the RIMs improvement in uncertainty modeling. In Component 1 (Fig. 10a), the RIMs is slightly more accurate than the baseline, its 95 % confidence interval is narrower and better aligned with the actual failure time. In contrast, the Naive SRUL exhibits a broader confidence interval and tends to overestimate RUL later in the test. Such overestimation can be risky in operational settings, as it may lead to unexpected failures and safety issues.

In Component 3 (Fig. 10b), the RIM outperforms the baseline in both accuracy and uncertainty. During the first half of the components lifetime, both methods track the ground truth reasonably well. However, in the second half, the Naive SRUL overestimates RUL and then experiences a sudden drop at failure, whereas the RIM mitigates this overestimation. This result highlights the RIMs strength in preserving accuracy while avoiding abrupt prediction collapses, which enhances its reliability for decision-making.

Table 4
Hyperparameters for the offline and online phase of the RIM using PF as base predictor.

Type	Learning rate	Iterations	Batch	k
Offline	1e-7	2000	True	-
Online	1e-5	100	True	3

3.3.2. *RIM with physics-guided PF as base predictor*

Similar to the HSMM approach, the physics-guided PF was trained using degradation histories from four single open-hole coupons. The training data were fitted to an exponential degradation model, yielding mean parameter values of $\epsilon_0 = 2.016 \times 10^{-6}$, $\beta = 0.1631$, and $\epsilon_\infty = 0.001409$. An example on the PF estimation of the observed strain data for coupon 5 is shown in Subfigure 11a.

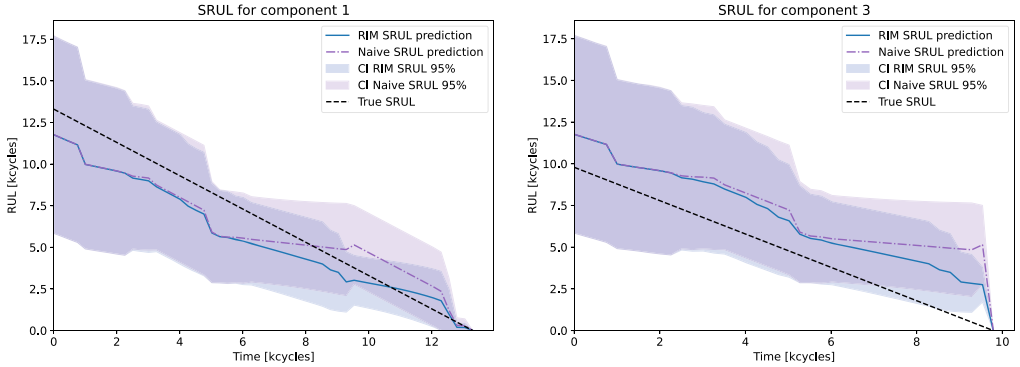
Using the coupon-level RUL as ground truth as reference, the average RMSE across 15 coupons was 8.01 kcycles. In general, the PF generated smooth coupon-level RUL trajectories (see Subfigure 11b), but its performance was more sensitive to noisy data than the HSMM. In particular, for some coupons in component 3, the exponential model struggled to capture late-life degradation, leading to divergence in the final predictions. As a result, some estimates failed to converge near the end of life, increasing errors (RMSE up to ~11 kcycles). This behavior indicates that while the PF is effective under stable conditions, it is less robust than the HSMM when degradation accelerates or sensor data becomes noisy.

The hyperparameters selected for the offline and online phases of the RIM with the physics-guided PF are summarized in Table 4.

Fig. 12 shows coupon-level RULs and the adapted RULs for components 1 and 3. In Component 1, the PF provided relatively stable predictions that the RIM successfully adapted into consistent component-level estimates. However, in Component 3, the coupon-level PF predictions became noisy and failed to converge near failure. Under these conditions, the RIM was still able to generate results, but the adapted RULs became unreliable close to the end of life. This emphasizes the need for diagnostic mechanisms to detect when input data quality is insufficient for reliable PF-based prognostics, as errors at the coupon level can otherwise propagate to the component level.

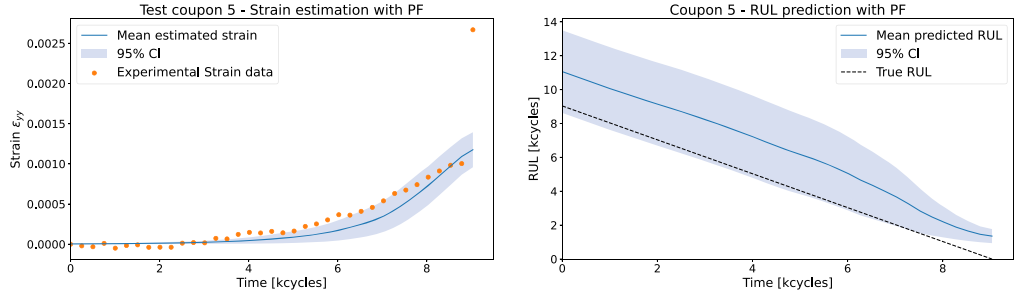
The overall performance comparison is reported in Table 5. The RIM improved RMSE in three out of four components, yielding a mean reduction from 7.90 to 7.32 kcycles. Although this reduction is smaller than with the HSMM, the improvement in uncertainty quantification was substantial: the CRPS dropped from 6.83 to 4.06, confirming that the RIM provides clearer and more reliable SRUL predictions even when coupon-level errors remain.

Table 6 presents the coverage and NMPIW results for the PF-based RIM. While coverage exhibits heterogeneous behavior across components, this must be interpreted jointly with the previously reported CRPS



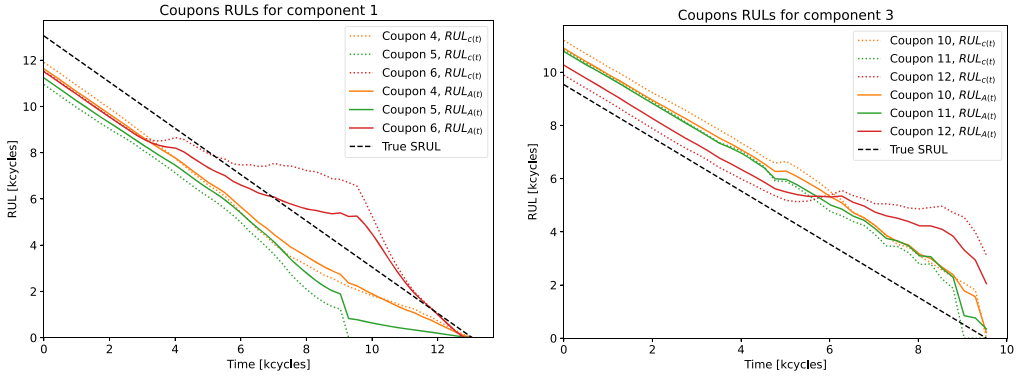
(a) Component 1. (b) Component 3.

Fig. 10. SRUL prediction using HSMM as base predictor for RIM vs Baseline with confidence intervals at 95 %.



(a) PF strain estimation. (b) PF RUL prediction.

Fig. 11. PF strain data estimation and RUL prediction at a coupon level.



(a) Component 1. (b) Component 3.

Fig. 12. Predicted component mean RULs and adapted mean RULs with RIM for each coupon for two components using PF as base predictor.

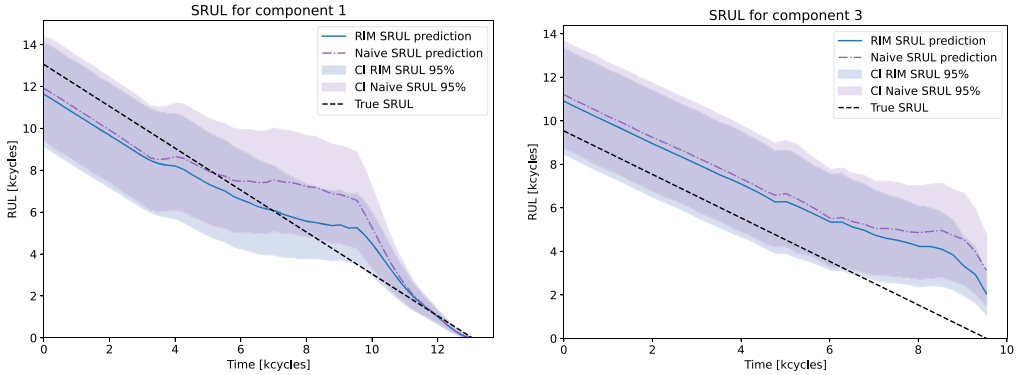
values in Table 5, which capture both predictive accuracy and uncertainty calibration. The RIM improves coverage for components 1-3, most notably for component 1 (from 0.83 to 0.94), while simultaneously reducing the CRPS across all components. This indicates that, even when coverage gains are modest, the predicted SRUL produced by the RIM is better aligned with the true SRUL.

For Component 4, the RIM leads to a decrease in coverage (0.85 to 0.45) and an increase in RMSE (6.67 to 7.89), indicating both reduced interval reliability and lower point accuracy. However, this is accompanied by a reduction in NMPIW (0.31 to 0.28) and an improvement in CRPS (5.40 to 4.97). Since CRPS jointly evaluates calibration and sharpness, this suggests that the RIM produces a more concentrated and overall better-aligned predictive distribution despite the degradation in point accuracy and empirical coverage. The results, therefore, reflect a shift from conservative, wider intervals to sharper predictions that are closer to the true SRUL distribution on average, albeit with some degree of undercoverage.

For the whole test set, the consistent reduction in mean CRPS, together with the decrease in NMPIW, indicates that RIM improves the sharpness-accuracy trade-off of the probabilistic SRUL predictions, even when coverage alone does not uniformly increase. These results highlight the importance of assessing uncertainty quality using complementary metrics rather than relying only on coverage.

Fig. 13 illustrates these results for Components 1 and 3. In Component 1, the RIM provides a more accurate mean RUL prediction than the baseline, along with a narrower 95 % confidence interval. By contrast, the Naive SRUL exhibits wider uncertainty and tends to overpredict the SRUL around the midpoint of the operation.

In Component 3, the RIM also outperforms the Naive SRUL. Although some coupon-level PF predictions did not converge, the RIM successfully



(a) Component 1. (b) Component 3.

Fig. 13. SRUL prediction using PF as base predictor for RIM vs Baseline with confidence intervals at 95 %.

Table 5
RMSE and CRPS for baseline (Naive SRUL) and RIM across four components using a PF as base predictor.

Component	RMSE [kcycles]		CRPS	
	Baseline	RIM	Baseline	RIM
1	5.53	**4.10**	4.34	**2.63**
2	9.75	**9.07**	8.42	**4.26**
3	9.65	**8.12**	9.16	**4.39**
4	**6.67**	7.89	5.40	**4.97**
Mean	7.90	**7.32**	6.83	**4.06**

Table 6
Coverage and NMPIW for baseline (Naive SRUL) and RIM across four components using a PF as base predictor.

Component	Coverage		NMPIW	
	Baseline	RIM	Baseline	RIM
1	0.83	0.94	0.31	0.29
2	0.47	0.50	0.47	0.44
3	0.64	0.69	0.49	0.46
4	0.85	0.45	0.31	0.28
Mean	0.70	0.65	0.40	0.37

adapted them to component-level predictions. While the mean RUL trajectory remains similar to the baseline, the RIM substantially improves uncertainty quantification, as reflected by the reduction in CRPS from 9.16 to 4.39.

Overall, the PF-based experiments reinforce the value of the RIM in reducing both error and uncertainty, even when the base predictor struggles with noisy CM data at the coupon level.

The case study highlights the distinct strengths of HSMM and PF as base predictors. The HSMM achieved lower RMSE and demonstrated robustness to noise, making it well suited for complex degradation patterns without requiring strong prior assumptions. The PF, while slightly less accurate, offered physics-based interpretability but showed greater sensitivity to noise and late-life deviations from its assumed degradation law.

In both cases, the RIM consistently improved accuracy and, more importantly, reduced CRPS, confirming its ability to enhance uncertainty quantification. This demonstrates the RIMs prognostic model-agnostic nature: it functions as a plug-and-play layer capable of adapting any base predictor into reliable component-level predictions.

4. Conclusions

This paper introduces the RUL Inoperabilities Model (RIM), a novel probabilistic framework inspired by the principles of the Inoperability Input-Output Model (IIM), to operate directly on coupon-level RUL predictions. The approach is model-agnostic, interpretable, and data-efficient, requiring only coupon-level training data and only one component-level degradation history for adaptation. By propagating uncertainty from the coupon level to the component level, the RIM provides not only point estimates of the System Remaining Useful Life (SRUL) but also well-calibrated confidence bounds, which are critical for decision-making in aerospace applications.

The methodology was validated through single open-hole aluminum coupon experiments using two distinct base predictors: a data-driven Hidden Semi-Markov Model (HSMM) and a physics-guided Particle Filter (PF). In both cases, the RIM consistently improved accuracy relative to a baseline Naive SRUL method, while providing superior uncertainty quantification. The CRPS values achieved with RIM were markedly lower than those of the baseline, demonstrating the frameworks ability to generate more reliable probabilistic predictions. Even when the PF struggled to capture late-life degradation behavior at the coupon level, the RIM adapted these imperfect predictions into SRUL predictions that more closely tracked the ground truth.

These findings demonstrate that the RIM can generalize effectively from limited training data, as it was trained offline with only one component, yet successfully tested on four unseen components. This capability is particularly important for aerospace structures, where full-scale experimental campaigns are costly and time-consuming. By showing consistent performance across different base predictors, the RIM establishes itself as a scalable solution for bridging the gap between coupon-level models and full-component prognostics.

Future work is needed to extend the RIM to capture epistemic uncertainties in the interdependency matrix, to integrate mechanisms for detecting unreliable coupon-level predictions before they propagate to the component level, and to validate the approach on larger and more complex aerospace assemblies under realistic loading and environmental conditions. Nevertheless, the present study demonstrates that the RIM provides a promising pathway toward practical, uncertainty-aware, and interpretable system-level prognostics.

CRediT authorship contribution statement

Mariana Salinas-Camus: Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing; **Oscar Carpentier:** Writing – original draft, Software, Methodology,

Investigation, Data curation, Conceptualization; **Nick Eleftheroglou:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

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.

References

[1] H.M. Elattar, H.K. Elminir, A.M. Riad, Prognostics: a literature review, Complex Intell. Syst. 2 (2) (2016) 125–154. https://doi.org/10.1007/s40747-016-0019-3

[2] M. Salinas-Camus, K. Goebel, N. Eleftheroglou, A comprehensive review and evaluation framework for data-driven prognostics: uncertainty, robustness, interpretability, and feasibility, Mech. Syst. Signal Process. 237 (2025) 113015. https://doi.org/10.1016/j.ymssp.2025.113015

[3] F. Tamssaouet, K.T.P. Nguyen, K. Medjaher, M.E. Orchard, System-level failure prognostics: literature review and main challenges, Proc. Inst. Mech. Eng. O 237 (3) (2023) 524–545. https://doi.org/10.1177/1748006X221118448

[4] S. Kim, J.-H. Choi, N.H. Kim, Challenges and opportunities of system-Level prognostics, Sensors 21 (22) (2021). https://doi.org/10.3390/s21227655

[5] A.A.R. Broer, R. Benedictus, D. Zarouchas, The need for multi-sensor data fusion in structural health monitoring of composite aircraft structures, Aerospace 9 (4) (2022). https://doi.org/10.3390/aerospace9040183

[6] H. Niu, J. Zeng, H. Shi, X. Zhang, J. Liang, G. Shi, Remaining useful life prediction for multi-component systems with stochastic correlation based on auxiliary particle filter, Reliab. Eng. Syst. Saf. 264 (2025) 111357. https://doi.org/10.1016/j.ress.2025.111357

[7] H. Khorasgani, G. Biswas, S. Sankararaman, Methodologies for system-level remaining useful life prediction, Reliab. Eng. Syst. Saf. 154 (2016) 8–18. https://doi.org/10.1016/j.ress.2016.05.006

[8] G. Galanopoulos, E. Fytsilis, N. Yue, A. Broer, D. Milanoski, D. Zarouchas, T. Loutas, A data driven methodology for upscaling remaining useful life predictions: from single- to multi-stiffened composite panels, Compos. C Open Access 11 (2023) 100366. https://doi.org/10.1016/j.jcomc.2023.100366

[9] L.R. Rodrigues, Remaining useful life prediction for multiple-component systems based on a system-level performance indicator, IEEE ASME Trans. Mechatron. 23 (1) (2018) 141–150. https://doi.org/10.1109/TMECH.2017.2713722

[10] K. Benaggoune, H. Mouss, A.A. Abdessemed, M. Bensakhria, Agent-based prognostic function for multicomponents system with stochastic dependence, in: 2018 International Conference on Applied Smart Systems (ICASS), 2018, pp. 1–5. https://doi.org/10.1109/ICASS.2018.8652027

[11] J. Wang, X. Wang, L. Wang, Modeling of BN lifetime prediction of a system based on integrated multi-Level information, Sensors 17 (9) (2017). https://doi.org/10.3390/s17092123

[12] A.R.-T. Palazuelos, E.L. Droguett, System-level prognostics and health management: a graph convolutional network-based framework, Proc. Inst. Mech. Eng. O 235 (1) (2021) 120–135. https://doi.org/10.1177/1748006X20935760

[13] A. Ifeanyi, J. Coble, Enhancing system-level prognostics with structural information: a graph-based approach, in: 2025 IEEE International Conference on Prognostics and Health Management (ICPHM), 2025, pp. 1–8. https://doi.org/10.1109/ICPHM65385.2025.11061821

[14] K.T.P. Nguyen, K. Medjaher, C. Gogu, Probabilistic deep learning methodology for uncertainty quantification of remaining useful lifetime of multi-component systems, Reliab. Eng. Syst. Saf. 222 (2022) 108383. https://doi.org/10.1016/j.ress.2022.108383

[15] J. Liu, E. Zio, System dynamic reliability assessment and failure prognostics, Reliab. Eng. Syst. Saf. 160 (2017) 21–36. https://doi.org/10.1016/j.ress.2016.12.003

[16] W.W. Leontief, Quantitative input and output relations in the economic systems of the united states, Rev. Econ. Stat. 18 (3) (1936) 105–125. https://doi.org/10.2307/1927837

[17] F. Tamssaouet, T.P.K. Nguyen, K. Medjaher, System-level prognostics based on inoperability input-output model, in: Annual Conference of the Prognostics and Health Management Society 2018, 2018, p. 0. https://doi.org/10.36001/phmconf.2018.v10i1.487

[18] F. Tamssaouet, K.T.P. Nguyen, K. Medjaher, M. Orchard, A fresh new look for system-level prognostics: handling multi-component interactions, mission profile impacts, and uncertainty quantification, Int. J. Progn. Health Manag. 12 (2) (2021). https://doi.org/10.36001/ijphm.2021.v12i2.2777

[19] T. Kontogiannis, M. Salinas-Camus, N. Eleftheroglou, Chapter ten - Hidden markov model applications: aviation prognostics, in: I.S. Triantafyllou, S. Malefaki, A. Karagrigoriou (Eds.), Stochastic Modeling and Statistical Methods, Advances in Reliability Science, Academic Press, 2025, pp. 191–213. https://doi.org/10.1016/B978-0-44-331694-4.00015-3

[20] M.D. Shields, K. Teferra, A. Hapij, R.P. Daddazio, Refined stratified sampling for efficient Monte Carlo based uncertainty quantification, Reliab. Eng. Syst. Saf. 142 (2015) 310–325. https://doi.org/10.1016/j.ress.2015.05.023

[21] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proc. IEEE 77 (2) (1989) 257–286. https://doi.org/10.1109/5.18626

[22] T. Loutas, A. Oikonomou, N. Eleftheroglou, F. Freeman, D. Zarouchas, Remaining useful life prognosis of aircraft brakes, Int. J. Progn. Health Manag. 13 (1) (2022). https://doi.org/10.36001/ijphm.2022.v13i1.3072

[23] N. Eleftheroglou, S.S. Mansouri, T. Loutas, P. Karvelis, G. Georgoulas, G. Nikolakopoulos, D. Zarouchas, Intelligent data-driven prognostic methodologies for the real-time remaining useful life until the end-of-discharge estimation of the Lithium-Polymer batteries of unmanned aerial vehicles with uncertainty quantification, Appl. Energy 254 (2019) 113677. https://doi.org/10.1016/j.apenergy.2019.113677

[24] T. Loutas, N. Eleftheroglou, G. Georgoulas, P. Loukopoulos, D. Mba, I. Bennett, Valve failure prognostics in reciprocating compressors utilizing temperature measurements, PCA-based data fusion, and probabilistic algorithms, IEEE Trans. Ind. Electron. 67 (6) (2019) 5022–5029. https://doi.org/10.1109/TIE.2019.2926048

[25] T. Loutas, N. Eleftheroglou, D. Zarouchas, A data-driven probabilistic framework towards the in-situ prognostics of fatigue life of composites based on acoustic emission data, Compos. Struct. 161 (2017) 522–529. https://doi.org/10.1016/j.compstruct.2016.10.109

[26] S. Xiaosheng, H. Li, Z. Zhang, N. Li, A Wiener-process-inspired semi-stochastic filtering approach for prognostics, Reliab. Eng. Syst. Saf. 249 (2024) 110200. https://doi.org/10.1016/j.ress.2024.110200

[27] N. Eleftheroglou, G. Galanopoulos, T. Loutas, Similarity learning hidden semi-Markov model for adaptive prognostics of composite structures, Reliab. Eng. Syst. Saf. 243 (2024) 109808. https://doi.org/10.1016/j.ress.2023.109808

[28] N.D. Alexopoulos, E. Migklis, A. Stylianos, D.P. Myriounis, Fatigue behavior of the aeronautical Al-Li (2198) aluminum alloy under constant amplitude loading, Int. J. Fatigue 56 (2013) 95–105. https://doi.org/10.1016/j.ijfatigue.2013.07.009

[29] R. Bai, Y. Li, J. Yin, Z. Jiao, K. Noman, Y. Wang, Physics-driven bayesian long short-term memory network for machinery remaining useful life prediction with uncertainty estimation, Reliab. Eng. Syst. Saf. 262 (2025) 111127. https://doi.org/10.1016/j.ress.2025.111127

[30] C. Jiang, T. Zhong, H. Choi, B.D. Youn, Physics-informed gaussian process probabilistic modeling with multi-source data for prognostics of degradation processes, Reliab. Eng. Syst. Saf. 258 (2025) 110893. https://doi.org/10.1016/j.ress.2025.110893

[31] H. Hersbach, Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather Forecast. 15 (5) (2000) 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

[32] A. Khosravi, S. Nahavandi, D. Creighton, A.F. Atiya, Lower upper bound estimation method for construction of neural network-based prediction intervals, IEEE Trans. Neural Netw. 22 (3) (2011) 337–346. https://doi.org/10.1109/TNN.2010.2096824

[33] J.I. Aizpurua, B.G. Stewart, S.D.J. McArthur, M. Penalba, M. Barrenetxea, E. Muxika, J.V. Ringwood, Probabilistic forecasting informed failure prognostics framework for improved RUL prediction under uncertainty: a transformer case study, Reliab. Eng. Syst. Saf. 226 (2022) 108676. https://doi.org/10.1016/j.ress.2022.108676

[34] T. Kontogiannis, M. Salinas-Camus, N. Eleftheroglou, HiMAP: hidden Markov for advanced prognostics, J. Open Source Softw. 11 (121) (2026) 9491. 10.21105/joss.09491. https://doi.org/10.21105/joss.09491

[35] M. Stone, Comments on model selection criteria of Akaike and Schwarz, J. Royal Stat. Soc. Ser. B (Methodol.) (1979) 276–278. https://doi.org/10.1111/j.2517-6161.1979.tb01084.x