Measurement Science and Technology

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Reliable prognosis for complex systems based on spatial-temporal representation modeling and uncertainty-aware dynamic Fusion

To cite this article: Zifei Xu *et al* 2026 *Meas. Sci. Technol.* **37** 245102

View the article online for updates and enhancements.

You may also like

- A Bayesian spatio-temporal degradation model based on ST-NaBGCN for aero-engine remaining useful life prediction and uncertainty quantification
Xia Wu, Zhiwen Liu and Lei Wang

- A remaining useful life prediction network using spatio-temporal soft threshold unit and adaptive multi-task weight allocation with missing sensor data
Weipeng Liu, Jingtian Shuai, Qidi Hu et al.

- Multiscale global and local self-attention-based network for remaining useful life prediction
Zhizheng Zhang, Wen Song, Qiqiang Li et al.

This content was downloaded from IP address 81.102.142.148 on 28/08/2026 at 10:10

Measurement Science and Technology

Check for updates

OPEN ACCESS

RECEIVED
27 March 2026

REVISED
16 May 2026

ACCEPTED FOR PUBLICATION
28 May 2026

PUBLISHED
15 June 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Reliable prognosis for complex systems based on spatial-temporal representation modeling and uncertainty-aware dynamic Fusion

Zifei Xu[1,2,3,*], **Musa Bashir**[1], **Qiang Zhang**[2,3], **Michael Beer**[1,4,5], **Jin Wang**[2] **and Zaili Yang**[2]

[1] Department of Civil and Environmental Engineering, University of Liverpool, Liverpool, United Kingdom
[2] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, United Kingdom
[3] School of Mechanical & Power and Energy Engineering, University of Shanghai for Science and Technology, Shanghai, People's Republic of China
[4] Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany
[5] International Joint Research Centre for Resilient Infrastructure & International Joint Research Centre for Engineering Reliability and Stochastic Mechanics,, Tongji University, Shanghai, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: zifei.xu@liverpool.ac.uk and z.xu@ljmu.ac.uk

Keywords: prognostics and health management, uncertainty quantification, information fusion, decision-making, Bayesian neural network

Abstract

Remaining useful life (RUL) prediction for complex systems is essential for ensuring their reliability and operational safety. Deep neural networks have demonstrated strong capabilities in prognostic modeling by learning degradation patterns from multivariate sensor data. However, their performance remains limited due to two key challenges: the difficulty of jointly capturing temporal degradation dynamics and cross-sensor dependencies, and the effective fusion of information from multiple variables. To overcome these challenges, this study develops an intelligent prognostic framework composed of three main modules: (i) a spatial–temporal feature embedding module for multivariable feature encoding; (ii) a Bayesian event-triggered dynamic graph update mechanism to robustly characterize evolving spatial relationships during degradation; and (iii) an uncertainty-aware fusion module for decision-level RUL prediction. Experiments on complex system RUL prediction tasks show that the proposed spatial–temporal network significantly improves prediction accuracy. The Bayesian event-triggered graph update further makes graph evolution more robust and better aligned with degradation progression. In addition, the uncertainty-aware fusion strategy improves uncertainty calibration and interval coverage, allowing a greater proporsion of true RUL values to fall within the predicted confidence bounds. The proposed framework make significant contribution on the provition of increased more reliable RUL prediction and improved support for predictive maintenance.

1. Introduction

1.1. Background

Remaining useful life (RUL) prediction is a cornerstone of prognostics and system health management (Xu *et al* 2023, Li *et al* 2026b). For relatively simple systems, physics-driven models have traditionally been used to estimate RUL. However, the increasing complexity of modern industrial systems, resulting from evolving operational requirements and market demands, poses significant challenges to purely physics-based RUL prediction (Lei 2016). In this context, data-driven approaches, especially neural network-based models, have become widely adopted because of their flexibility, capacity to handle large-scale data, and strong ability to learn complex degradation patterns from monitoring signals (Lei *et al* 2022, Yang *et al* 2025).

Data-driven RUL prediction models, particularly those based on neural networks, have achieved notable success across a wide range of engineering applications. Despite these encouraging advances

© 2026 The Author(s). Published by IOP Publishing Ltd

in RUL prediction for complex systems such as batteries (Liu *et al* 2021) and aircraft (Meng *et al* 2023, Karatzinis *et al* 2024, Liu *et al* 2024, Guo *et al* 2025), comparatively little attention has been paid to the trustworthiness of neural network outputs in real-world decision-making contexts, where predictions directly influence maintenance scheduling, safety assurance, and cost–risk trade-offs (Ovadia *et al* 2019, Zhou *et al* 2023, Xu *et al* 2025). Consequently, it remains crucial to understand not only how accurate a neural network–based RUL model can be, but also how reliable, robust, and uncertainty-aware its predictions are under varying operating conditions and previously unseen degradation patterns (Laflamme *et al* 2026).

To position this study within the broader research landscape, section 1.2 critically reviews recent advances in RUL prediction, with a particular emphasis on neural network–based paradigms for learning and evaluation.

1.2. State of the art

In neural network–driven RUL prediction, a common underlying assumption is that a sufficiently expressive model exists that can directly map monitoring data to RUL, despite the complex, stochastic, and non-stationary nature of degradation processes (Ni *et al* 2022, Feng *et al* 2023a). From a retrospective viewpoint, such a neural network–based model effectively acts as a surrogate inference model, which estimates latent parameters based on observed data. In this context, the parameter of interest is the RUL, while the observations correspond to condition monitoring data collected from the system (Shi *et al* 2025).

In this literature review, neural network–based RUL prediction approaches are broadly categorized into deterministic prediction models and uncertainty-aware (probabilistic) prediction models.

For deterministic RUL prediction, a typical RUL prediction framework consists of two phases: an offline training phase, during which the model is learned using historical data, and an online inference phase, in which the trained model is deployed in industrial applications to estimate RUL based on incoming condition monitoring data. During the training process, the primary focus is on learning a model that can accurately map condition monitoring data to the corresponding condition labels, commonly referred to as RUL labels. From a technical perspective, such a model typically consists of an encoder, which extracts representative information from raw monitoring data, and a decoder, which maps the learned representation to RUL estimates.

Here, the function of the encoder is critical to the generalization capability of neural network–based RUL prediction models. In practice, most RUL prediction tasks focus on systems exhibiting gradual degradation processes, which naturally motivates the design of encoders that emphasize temporal modeling of condition monitoring data. Consequently, a wide range of time-series encoders, including causal convolutional networks, recurrent neural networks and their variants, as well as Transformer-based architectures, have been adopted to capture temporal degradation patterns. While these temporal encoders have demonstrated effectiveness in modeling degradation evolution over time, they primarily focus on temporal dependencies and often treat multivariate sensor signals in a spatially homogeneous or implicit manner. As a result, the explicit modeling of spatial relationships among sensors or system components remains underexplored, motivating recent efforts toward incorporating spatial and spatio-temporal (ST) encoding mechanisms for more informative and robust RUL prediction.

Besides, deterministic RUL prediction research has also emphasized scenario adaptation, especially when full life-cycle degradation samples or RUL labels are difficult to obtain. These challenges are typically addressed through three major directions: small-sample learning, where limited degradation data are available (Zhu *et al* 2022, Yin *et al* 2026); For instance, an adaptive sampling model was developed to reconstruct small simples, representing a novel solution for failure information extraction under limited data (Li *et al* 2026a). Robust learning under noisy or incomplete measurements (Hervé de Beaulieu *et al* 2024, Nejjar *et al* 2024, Zhao *et al* 2026); and domain adaptation or transfer learning, which aims to generalize RUL models across different operating conditions or system configurations (Du *et al* 2024, Le Xuan *et al* 2024, Lu *et al* 2024, Nejjar *et al* 2024, Li *et al* 2025).

However, the aforementioned studies largely overlook the impact of uncertainty on the reliability of RUL prediction (Feng *et al* 2023b). This includes not only how data uncertainty, arising from measurement noise, operational variability, and incomplete degradation observations, is propagated through AI models and quantified at the output level, but also how model uncertainty can be characterized to reflect the confidence of the learned model itself when facing previously unseen conditions or degradation patterns.

In probabilistic RUL prediction, aleatoric uncertainty arising from the inherent randomness of degradation and failure processes is often modeled using physics- or reliability-inspired approaches, such as stochastic degradation models and survival analysis (Pan *et al* 2024, Zhang *et al* 2024). Compared with neural networks that directly output predictive distributions, stochastic degradation and survival models offer physically interpretable formulations of aleatoric uncertainty but are constrained by strong modeling assumptions and limited flexibility in handling high-dimensional, non-stationary degradation data. In contrast, distributional neural networks provide greater modeling flexibility at the cost of reduced interpretability and reliance on data-driven calibration (Han and Li 2022, Amin *et al* 2023, Zhang *et al* 2025). While aleatoric uncertainty characterizes the inherent randomness of degradation and failure processes and captures how uncertainty propagates from system evolution to RUL, epistemic uncertainty reflects the model's confidence in its predictions given limited data and imperfect knowledge. In the context of RUL prediction, where prognostic results directly inform risk-aware maintenance and operational decision-making, the modeling and quantification of epistemic uncertainty are therefore of critical importance, as they provide an objective measure of the reliability of model predictions and support informed decision-making under uncertainty (Peng *et al* 2020, Sedehi *et al* 2022, Costa *et al* 2024, Mukhoti *et al* 2024, Tian *et al* 2026). Practical cases demonstrate that novel models have been developed to quantify multiple types of uncertainties, contributing to strengthened understanding in addressing uncertainties in reliability, risk, and maintenance investigations (Li *et al* 2020, 2021). However, the reliability and trustworthiness of such uncertainty quantification methods remain an open question. Typically, uncertainty in neural network–based RUL prediction is quantified using Monte Carlo (MC) dropout (Nguyen *et al* 2022), Bayesian neural networks (Cheng *et al* 2024, Keshun *et al* 2024), and ensemble learning (Soualhi *et al* 2023), which primarily capture epistemic uncertainty related to model confidence (Zhan *et al* 2024). The above studies demonstrate that incorporating prediction uncertainty into neural network-based approaches is essential for reliable RUL prediction.

1.3. Research gaps and contribution

Based on the above state-of-the-art review, two key requirements can be identified for reliable RUL prediction in complex industrial systems. First, existing studies need to move beyond deterministic prediction toward uncertainty-aware prognostics, so that the model can provide not only point estimates but also reliable confidence information for risk-sensitive decision-making. Second, for multivariate sensor-based systems, it remains essential to learn informative degradation representations from high-dimensional observations, rather than being misled by noise or spurious inter-sensor dependencies caused by sensor variability during degradation. These challenges indicate that predictive accuracy and uncertainty quantification should be addressed in a unified framework rather than being treated separately.

Motivated by these considerations, this work proposes an uncertainty-aware spatial–temporal prognostic framework for multivariate RUL prediction. The framework integrates a dedicated embedding module with a dynamic sensor-graph learning mechanism, in which the adjacency matrix is updated online according to learned events in the latent space, allowing inter-sensor relationships to evolve with system degradation. In this way, the model is able to learn task-relevant spatial–temporal degradation representations that better characterize the progression of failure modes. Furthermore, an uncertainty-aware multi-sensor decision fusion scheme is introduced to aggregate per-sensor predictions according to their estimated uncertainties, thereby improving both prediction robustness and reliability.

Overall, the proposed framework is designed to provide a more trustworthy basis for RUL prognostics and to support risk-aware maintenance decision-making. The main contributions of this work are summarized as follows:

1. An uncertainty-aware spatial–temporal prognostic model for multivariate RUL prediction was developed. The proposed model is capable of learning informative spatial–temporal degradation representations from multivariate sensor data by capturing both temporal evolution patterns and cross-sensor dependencies. In addition, it explicitly models heterogeneous uncertainty by jointly considering aleatoric uncertainty and epistemic uncertainty, thereby providing a more reliable basis for prognostic prediction.
2. A posterior-motivated event-triggered dynamic graph update mechanism was proposed. Instead of relearning a fully unconstrained graph at every time step, the method starts from a base graph derived from sensor embeddings and updates it selectively when the triggering criterion suggests a meaningful change in inter-sensor relationships. This design reduces unnecessary graph fluctuations and provides a more structured way to analyze dependency evolution during degradation.
3. An uncertainty-aware fusion strategy for RUL estimation was proposed. The model estimates sensor-level predictive uncertainty and uses it to modulate the contribution of individual sensors to the final

prediction. The performance of the fusion strategy was evaluated on prediction robustness under noisy or inconsistent sensor conditions.

The remainder of the paper is structured as follows. Section 2 presents the preliminary methods used in the proposed approach. Section 3 introduces the proposed methodologies, section 4 presents the case study and discussion. Section 5 concludes the study, while discussing future directions for reliable intelligent prognosis.

2. Preliminaries

2.1. Problem definition

From a statistical learning perspective, formulating the RUL prediction in terms of the conditional distribution $\wp_\theta(\mathbb{Y}|\mathbb{X})$ implies that the model aims to learn not only a point estimate of the RUL target, but the entire stochastic relationship between sensor observations and the corresponding RUL. In this formulation, the target variable $\mathbb{Y}$ is treated as a random variable conditioned on the input $\mathbb{X}$, rather than as a determinstic function of the measurement (Qiu *et al* 2025).

Under this assumption, the baseline model is trained by minimizing a data-driven learning risk defined as the expected negative log-likelihood (NLL) of the conditional prespective distribution:

$$\mathbb{R}_{\text{data}}(\theta) = \mathbb{E}_{(\mathbb{X},\mathbb{Y})}\left[-log\wp_\theta(\mathbb{Y}|\mathbb{X})\right] \tag{1}$$

which corresponds to the expected NLL risk in statistical learning. For continuous degradation variables, characterizing a conditional distribution inherently requires the specification of both a central tendency and a measure of dispersion. Accordingly, the predictive distribution is parameterized by an input-dependent mean and variance, and is modeled as:

$$\wp_\theta(\mathbb{Y}|\mathbb{X}) = \mathcal{N}\left(\mu_\theta(\mathbb{X}), \sigma^2{}_\theta(\mathbb{X})\right) \tag{2}$$

where $\mu_\theta(\mathbb{X})$ represents the predicted mean RUL and $\sigma^2{}_\theta(\mathbb{X})$ captures the aleatoric uncertainty arising from stochastic operating conditions and measurement noise. Under this Gaussian assumption, minimizing the expected NLL risk is equivalent to minimizing the expected loss:

$$\mathcal{L}_{\text{data}} = \mathbb{E}_{(\mathbb{X},\mathbb{Y})}\left[\frac{1}{2}\left(\log\sigma^2{}_\theta(\mathbb{X}) + \frac{(\mathbb{Y}-\mu_\theta(\mathbb{X}))^2}{\sigma^2{}_\theta(\mathbb{X})}\right)\right]. \tag{3}$$

In this formulation, the squared error term enforces consistency between the predicted mean and the observed degradation targets, while the logarithmic variance term penalizes degenerate uncertainty estimates. As a result, the heteroscedastic variance acts as an adaptive weighting mechanism on the learning risk, allowing observations associated with higher intrinsic noise to exert a reduced influence on parameter optimization. This baseline objective therefore realizes a principled form of empirical risk minimization under a probabilistic framework, providing a statistically grounded foundation for RUL prediction in the presence of heterogeneous degradation patterns and stochastic measurement effects.

2.2. Graph-structured long short-term memory network

In the multi-sensor monitoring setting for RUL prediction, a graph-structured long short-term memory network is adopted (GLSTM) as the backbone, which augments a standard LSTM with a learnable sensor interaction graph so that each sensor's hidden state is updated by jointly aggregating temporal information from its own history and spatial information propagated from other correlated sensors. Formally, the sensor measurements are represented as:

$$\mathbb{X} = \begin{bmatrix} \mathbb{X}_1^{\mathrm{T}} \\ \vdots \\ \mathbb{X}_T^{\mathrm{T}} \end{bmatrix}, \mathbb{X}_T = \left[x_{t,1}, x_{t,2}, \ldots, x_{t,N}\right]^{\mathrm{T}} \tag{4}$$

where $x_{t,i}$ is the scalar measurement of the ith sensor at time t. Before feeding the data into recurrent backbone, each scalar sensor signal is mapped into a latent space and augmented with sensor-positional bias. Specifically, for each sensor i at time step t, defined as:

$$u_{t,i} = \phi(x_{t,i}) \in \Re^K \tag{5}$$

where $\phi(\cdot)$ is a shared nonlinear mapping implemented as a small feed-forward network with a linear layer, activation, and normalization; K is the embedding dimension. In addition, each sensor index i is associated with a learnable positional embdding $\mathscr{b}_i \in \Re^K, i = 1, 2, \ldots, N$, which captures sensor-specific biases and identity-related characteristics independent of the instantaneous measurement. Therefore, the final input feature of sensor i at time t is given by equation (6), which serves as the input to the GLSTM.

$$\mathfrak{z}_{t,i} = \mathfrak{u}_{t,i} + \mathscr{b}_i. \tag{6}$$

For all sensor, input node features are obtained by stacking all sensors at time t:

$$\mathbb{z}_t = \begin{bmatrix} z_{t,1} \\ \vdots \\ z_{t,N} \end{bmatrix} \in \Re^{N\times K} \tag{7}$$

$$\mathbb{A}_{ij} = \exp\left(\frac{\mathfrak{s}_{ij}}{\tau}\right) / \sum_{N}^{k=1} \exp\left(\frac{\mathfrak{s}_{ik}}{\tau}\right). \tag{8}$$

In GLSTM, a learnable sensor interaction graph is used to model spatial dependencies among sensors, each sensor with graph embedding computed by The parwise similarity between sensor i and j: $\&_{ij} = g_i^T g_j$. These similarities are converted into row-normalized adjacency matrix $\mathbb{A} \in \Re^{N\times N}$ through a softmax with a temperature parameter $\tau < 0$.

The resulting graph $\mathcal{G}$ is denoted by equation (9)

$$\mathbb{G} = (\mathbb{V}, \mathbb{A}), \mathbb{V} = \{1, 2, \ldots, N\}. \tag{9}$$

Given the input node embedding features $\mathbb{Z}_t$ and graph $\mathbb{A}$, the GLSTM updates the hidden and cell states of all sensors in a ST manner. Let $\mathbb{H}_{t-1}$ and $\mathbb{C}_{t-1}$ denote the hidden and cell state of the GLSTM at time $t-1$. At each time step, there is a transformation by adjacency matrix $\mathbb{A}$:

$$\widetilde{\mathbb{H}}_{t-1} = \mathbb{A} \cdot \mathbb{H}_{t-1} \in \Re^{\mathrm{N}\times\mathrm{D}} \tag{10}$$

where D is the hidden dimension in LSTM. In addition to the static adjacency martic $\mathbb{A}_{ij}$, dynamic adjacency (DA) is also considered to capture time-varying sensor interactions, The pairwise similarity is extended to a time-dependent form: $\&^{(t)}{}_{ij} = g^{(t)}{}_i^{\mathrm{T}} g^{(t)}{}_j$, where $g^{(t)}{}_i$ denotes the embedding of sensor i at time step t. The DA matrix is then obtained via the same row-wise softmax normalization as in equation (11):

$$\mathbb{A}^{(t)}{}_{ij} = \exp\left(\frac{\mathfrak{s}^{(t)}{}_{ij}}{\tau}\right) / \sum_{k=1}^{N} \exp\left(\frac{\mathfrak{s}^{(t)}{}_{ik}}{\tau}\right). \tag{11}$$

The resulting time-dependent graph is denoted as:

$$\mathbb{G} = \left(\mathbb{V}, \mathbb{A}^{(t)}\right), \ \mathbb{V} = \{1, 2, \ldots, N\}. \tag{12}$$

Accordingly, the spatial transformation in equation (10) is extended to the dynamic case as:

$$\widetilde{\mathbb{H}}_{t-1} = \mathbb{A}^{(t)} \cdot \mathbb{H}_{t-1} \in \Re^{\mathrm{N}\times\mathrm{D}}. \tag{13}$$

As a result, each sensor node receives information from other correlated sensors according to the learned adjacency. The GLSTM gates are then computed by combing $\mathbb{Z}_t$ and the hiddens:

$$\begin{bmatrix} \mathbb{i}_t \\ \mathbb{f}_t \\ \mathbb{o}_t \\ \mathbb{g}_t \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{bmatrix} \left(\mathbb{Z}_t \mathbb{W}_z + \widetilde{\mathbb{H}}_{t-1} \mathbb{W}_h + \mathbb{b}\right) \tag{14}$$

$$\mathbb{C}_t = \mathbb{f}_t \odot \mathbb{C}_{t-1} + \mathbb{i}_t \odot \mathbb{g}_t \tag{15}$$

$$\mathbb{H}_t = \mathbb{O}_t \odot \tanh(\mathbb{C}_t) \tag{16}$$

where $\mathbb{i}_t\ \mathbb{f}_t\ \mathbb{o}_t\ \mathbb{g}_t \in \Re^{N\times D}$ are the input, forget, output and candidate gates respectively; $\mathbb{W}_z, \mathbb{W}_h \in \Re^{K\times D}$ and $\mathrm{b} \in \Re^{D}$ are learnable parameters; σ is the sigmoid function applied element-wise.; $\odot$ denotes element-wise multiplication. The resulting $\mathbb{H}_t$ encodes the ST health representation of all sensors at time t, which is subsequently aggregated and fed into the prediction head to estimate the RUL.

2.3. Probabilistic neural network for uncertainty capture

In RUL estimation, uncertainty arises not only from stochastic measurement noise and operational variability, but also from limited data coverage and model uncertainty. To explicitly characterize these two sources, the proposed network predicts both the RUL mean and an input-dependent variance. For an input sequence $\mathbb{X}$, the network parameterized by θ defines the conditional predictive distribution as:

$$\mathbb{Y}|\mathbb{X},\,\theta \sim \mathcal{N}\left(\mu_\theta\left(\mathbb{X}\right),\,\sigma_\theta^2\left(\mathbb{X}\right)\right) \tag{17}$$

where $\mu_\theta\left(\mathbb{X}\right)$ is the predicted RUL mean and $\sigma_\theta^2\left(\mathbb{X}\right)$ is the heteroscedastic noise variance estimated by the probabilistic head. The variance term models the aleatoric uncertainty, i.e. the irreducible uncertainty caused by measurement noise and stochastic variability in the degreadation process.

From a Bayesian perspective, the network parameters are treated as random variables as well with posteriro distribution $\wp(\theta|D)$, where $\mathfrak{D} = \{\mathbb{X},\,\mathbb{Y}\}$ demotes the training dataset. Accordingly, the Bayesian predictive distribution is written as:

$$\mathfrak{D}\left(\mathbb{Y}|\mathbb{X},\mathbb{D}\right) = \int \mathfrak{D}\left(\mathbb{Y}|\mathbb{X},\boldsymbol{\theta}\right)\mathfrak{D}\left(\boldsymbol{\theta}|\mathfrak{D}\right)\mathfrak{d}\boldsymbol{\theta}. \tag{18}$$

For deep neural networks, the integral in equation (18) is generally intractable. So in this paper, MC dropout is adopted to approximate the posterior over model parameters. By keeping dropout active at test time, each stochastic forward pass corresponds to sampling an approximate parameter realization $\boldsymbol{\theta}^{(k)} \sim q\left(\boldsymbol{\theta}\right)$, where $q\left(\boldsymbol{\theta}\right)$ is the dropout-induced variational approximation to $\mathfrak{D}(\boldsymbol{\theta}|\mathfrak{D})$. For the kth stochastic pass, the network outputs a Gaussian predictive component with mean $\mu^{(k)}\left(\mathbb{X}\right)$ and $\sigma^{2(k)}\left(\mathbb{X}\right)$.

$$\mathfrak{D}\left(\mathbb{Y}|\mathbb{X},\boldsymbol{\theta}^{(k)}\right) = \mathcal{N}\left(\mathbb{Y};\mu^{(k)}\left(\mathbb{X}\right),\sigma^{2(k)}\left(\mathbb{X}\right)\right). \tag{19}$$

Using K stochastic forward passes, the predictive mean is approximated as:

$$\hat{\boldsymbol{\mu}}_t = \frac{1}{K}\sum_{K}^{k=1}\mu_t^{(k)}. \tag{20}$$

According to the law of total variance, the predictive variance

$$\mathbb{A}_{ij}^{(t)} = \exp\left(\frac{\mathfrak{s}_{ij}^{(t)}}{\tau}\right) \Big/ \sum_{k=1}^{N}\exp\left(\frac{\mathfrak{s}_{ik}^{(t)}}{\tau}\right) \tag{11}$$

can be decomposed as:

$$\mathrm{Var}(Y\mid X,\mathcal{D}) = \mathbb{E}_{p(\theta|\mathcal{D})}[\mathrm{Var}(Y\mid X,\theta)] + \mathrm{Var}_{p(\theta|\mathcal{D})}[\mathbb{E}(Y\mid X,\theta)] \tag{20}$$

Substituting $\mathfrak{E}\left(\mathbb{Y}|\mathbb{X},\boldsymbol{\theta}\right) = \mu_\theta\left(\mathbb{X}\right)$ and $\mathrm{Var}\left(y|\mathbb{X},\boldsymbol{\theta}\right) = \sigma_\theta^2\left(\mathbb{X}\right)$ into the above expression shows that the first term corresponds to the expected data noise level, whereas the second term measures the variability induced by parameter uncertainty. Its MC approximation is given by:

$$\widehat{Var}\left(\left(\mathbb{Y}|\mathbb{X},D\right)\right) \approx \underbrace{\frac{1}{K}\sum_{k=1}^{K}\sigma^{2(k)}\left(\mathbb{X}\right)}_{aleatoric} + \underbrace{\frac{1}{K}\sum_{k=1}^{K}\left(\mu^{(k)}\left(\mathbb{X}\right) - \hat{\boldsymbol{\mu}}\left(\mathbb{X}\right)\right)^2}_{epistemic} \tag{22}$$

where the first term represents aleatoric uncertainty learned by the heteroscedastic probabilistic head, and the second term represents epistemic uncertainty quantified by the variability across stochastic dropout predictions. In the proposed framework, both uncertainty components are further exploited by the uncertainty-aware dynamic fusion (UDF) module described in section 3.3.

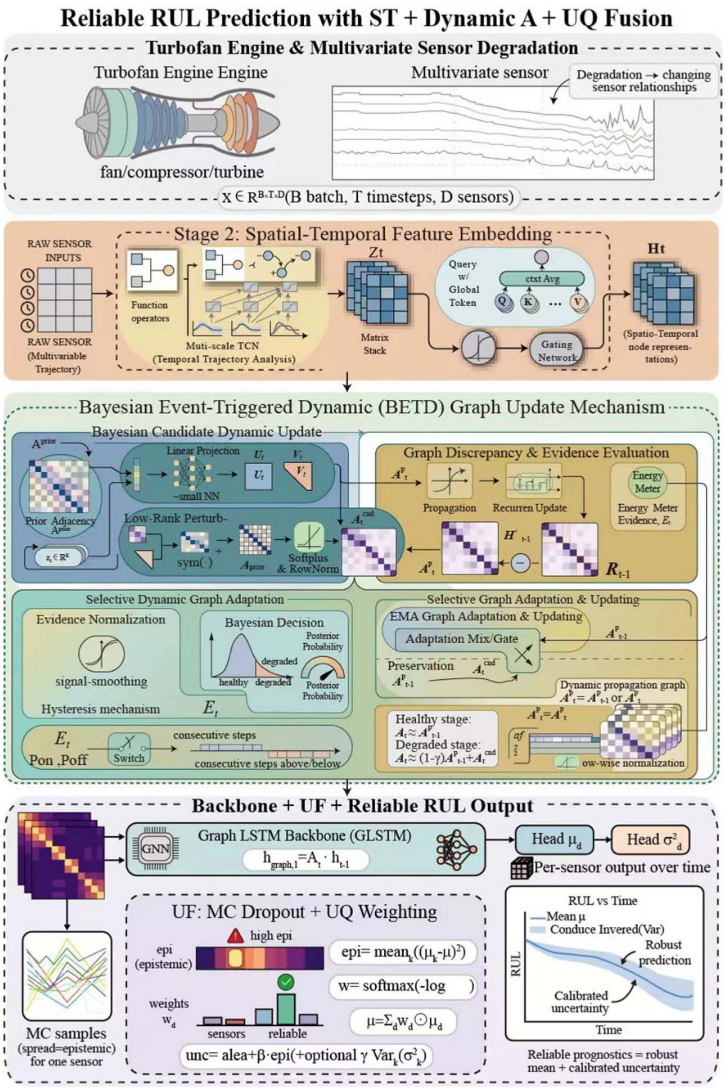


Figure 1. The proposed prognostic framework.

3. Methodology

3.1. Overall framework for reliable prognosis

Figure 1 summarizes the proposed prognostic framework is a unified reliable RUL prediction architecture for multivariate degradation signals.

It consists of four key components: spatial–temporal feature embedding, Bayesian event-triggered dynamic (BETD) graph updating, graph-based temporal backbone modeling, and uncertainty-aware output fusion. Starting from raw multi-sensor trajectories, the framework first extracts informative degradation representations by jointly learning temporal evolution and inter-sensor dependencies. It then adaptively updates the graph structure to reflect changing sensor relationships during degradation. Based on the updated graph, a graph-temporal backbone is used to model the degradation process and estimate RUL. Finally, uncertainty quantification and fusion are incorporated to produce calibrated and robust RUL predictions with improved reliability.

3.2. ST feature embedding

A ST feature embedding module is developed, which injects explicit inductive biases into per-sensor representations through multi-scale temporal modeling and sensor-wise attention–based interaction learning.

To address this, the ST embedding layer is designed. It first produces structured, noise-aware node representations for graph modeling, so that subsequent graph-structured recurrent propagation operates on semantically enriched and well-conditioned features rather than raw sensor inputs. The ST strucutral diagram is shown in figure 2.

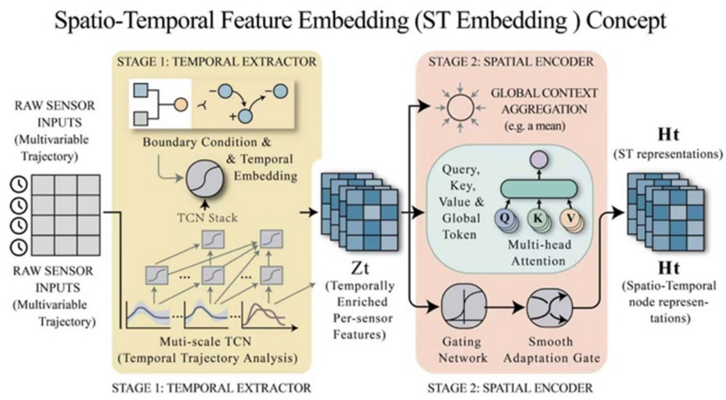


Figure 2. Spatial-temporal feature embedding module.

In the ST embedding module, it can update the simple per-sensor mappling in equation (5) by two-stage spatio-temporal encoder. At time step t, first and second difference channels for the multivariable are constructed as follows:

$$\Delta\mathbb{X}_t = \mathbb{X}_t - \mathbb{X}_{t-1}, \Delta^2\mathbb{X}_t = \Delta\mathbb{X}_t - \Delta\mathbb{X}_{t-1} \tag{21}$$

where zero padding for boundary conditions, i.e. $\mathbb{X}_t = 0$ for $t \leqslant 0$. Stacking the three channels yields descriptor matrix:

$$\mathbb{S}_t = \left[\mathbb{X}_t, \Delta\mathbb{X}_t, \Delta^2\mathbb{X}_t\right]. \tag{22}$$

This descriptors are first mapped into a K-dimentional latent feature space by shared linear projection:

$$\mathbb{H}_t^{(0)} = \mathbb{S}_t\mathbb{W}^{(0)} + \mathbb{b}^{(0)} \tag{23}$$

where $\mathbb{W}^{(0)} \in \Re^{3\times K}$ and $\mathrm{b}^{(0)} \in \Re^K$ are learnable parameters. The bias will broadcasting to all sensors. The ith row,denoted by $\mathrm{h}_{t,i}^{(0)} \in \Re^K$ is thus the initial temporal embedding of sensor i at time t.

On top of $\{\mathbb{H}_t^{(0)}\}_{t=1}^{\mathrm{T}}$, a stack of L causal dilated 1D convolutions is applied along the time axis. For ℓth layer, the update for sensor i can be weitten in standard convolution form as:

$$\mathbb{h}_{t,i}^{(\ell)} = \mathrm{SiLU}\left(\left(\mathbb{h}_i^{(\ell-1)} * \mathbb{W}^{(\ell)}\right)_t + \mathbb{b}^{(\ell)}\right) \tag{24}$$

where $\mathbb{h}_{t,i}^{(\ell)}$ is the temporal temporal trajectory of sensor i at layer $\ell-1$; $\mathbb{W}^{(\ell)} = \left\{\mathbb{W}_k^{(\ell)}\right\}_{k=0}^{k=\mathcal{K}_c-1}$ is the convolution kernel, $\mathbb{W}_k^{(\ell)} \in \Re^{K\times K}$, $\mathbb{b}^{(\ell)} \in \Re^K$ is a bias vector; $\mathrm{SiLU}(\cdot)$ is the activation. Particually, $*$ demotes dilated causal convolution as:

$$\left(\mathbb{h}_i^{(\ell-1)} * W^{(\ell)}\right)_t = \sum_{k=0}^{\mathcal{K}_c-1} \mathrm{SiLU}\left(\mathbb{W}_k^{(\ell)} \cdot \mathbb{h}_{t-\mathfrak{o}\cdot k,i}^{(\ell-1)}\right),$$

$$\mathbb{h}_{\tau,i}^{(\ell-1)} \equiv 0 \; for \; \tau \leqslant 0 \tag{25}$$

which is the standard 1D convolution with kernel size $\mathcal{K}_c$, dilation $\mathfrak{d}$ and causal padding. For notational compactness, all sensors into the array form $\mathbb{H}_t^{(\ell)} \in \Re^{N\times K}$ by stacking $\mathrm{h}_{t,i}^{(\ell)}$ are collected. The output of the temporal stack $\{\mathbb{H}_t^{(L)}\}_{t=1}^{T}$ serves as the temporally enriched per-sensor features that are further processed by the spatial encoder.

A global context vector is first aggregated by averaging over all sensor features and passing it through a small feed-forward network:

$$\tilde{\mathbb{h}}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbb{h}_{t,i}^{(L)}, \mathbb{h}_t = \psi\left(\tilde{\mathbb{h}}_t\right). \tag{26}$$

Where $\psi(\cdot)$ denotes a two-layer MLP followed by layer normalization. $\mathbb{h}_{t,i}^{(L)}$ is the temporally enriched feature of sensor i at time t. The spatial encoder refines $\mathbb{h}_{t,i}^{(L)}$ by explicitly modeling cross-sensor interactions via attention with a learnable global token. The global token $\mathbb{g}_t$ summarizes the overall system state at time t.

The query, key, and value matrices for multi-head attention are then constructed as follows:

$$\mathcal{Q}_t = \mathrm{LN}\left(\mathbb{H}_t^{(L)}\right),$$

$$\mathbb{K}_t = \mathbb{V}_t = LN\left(\left[\begin{array}{c} \mathbb{g}_t{}^{\mathrm{T}} \\ \left(\mathbb{H}_t^{(L)}\right)^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}\right) \tag{27}$$

where $\mathrm{LN}(\cdot)$ is layer normalization operators. Multi-head self-attention with the global token is then given by:

$$\tilde{\mathbb{H}}_t = \mathbb{H}_t^{(L)} + Atten\left(\mathcal{Q}_t, \mathbb{K}_t, \mathbb{V}_t\right) \tag{28}$$

where $Atten(\cdot)$ denotes the standard multi-head attention operator with residual connection:

$$\tilde{\mathbb{H}}_t = \mathrm{LN}\left(\tilde{\mathbb{H}}_t + \mathrm{FFN}\left(\tilde{\mathbb{H}}_t\right)\right) \tag{29}$$

where $\mathrm{FFN}(\cdot)$ is a two-layer MLP applied row-wise. To obtain sensor-wise correction signals, a broadcasted copy of the global token is concatenated and passed through a gating network as follows:

$$\mathbb{G}_t = \sigma\left(\mathbb{W}_2 SiLU\left(\mathbb{W}_1\left[\hat{\mathbb{H}}_t || \mathbb{g}_t{}^T\right] + \mathbb{b}_1\right) + \mathbb{b}_2\right) \tag{30}$$

where $[\cdot||\cdot]$ denotes channel-wise concatenation, $\mathbb{W}_1, \mathbb{W}_2, \mathbb{b}_1, \mathbb{b}_2$ are learnable parameters; $\sigma(\cdot)$ is the sigmoid function. he gated correction and the final ST embedding are then defined as:

$$\begin{aligned} \Delta\mathbb{H}_t &= \mathbb{G}_t \odot \widehat{\mathbb{H}}_t \\ \mathbb{Z}_t &= LN\left(\mathbb{U}_t + \mathbb{H}_t^{(L)} + \Delta\mathbb{H}_t\right) \end{aligned} \tag{31}$$

where $\odot$ is the Hadamard product. $\mathbb{U}_t$ stacks the per-sensor embeddings from equation (5). The matrix $\mathbb{Z}_t$ thus constitutes the ST node representations at time t, which are subsequently fed into the GLSTM backbone for graph-structured degra-dation modeling.

3.3. BETD graph update

In the baseline GLSTM, the spatial dependency among sensors is modeled by a fixed adjacency matrix $\mathbb{A}^{\mathrm{prior}}$ as defined in equation (8), which remains constant over time. Although this formulation ensures structural stability, it cannot capture the evolving inter-sensor relationships that emerge during system degradation. A straightforward extension is to update the adjacency matrix at every time step as a function of the current observation, which can be generally written as: $\mathbb{A}^{(t)} = f(\mathbb{Z}_t)$, where $\mathbb{Z}_t$ denotes the node embeddings at time step t. However, updating the adjacency matrix at every time step may cause unnecessary structural fluctuations, despite the gradual nature of degradation-induced coupling changes.

To address this issue, a BETD graph update mechanism is proposed, as illustrated in figure 3. Instead of fully replacing the prior adjacency at every time step, BETD first constructs a candidate dynamic graph from the current context, and then updates the graph only when sufficient evidence indicates that the current relational structure has meaningfully changed. In this way, the graph remains stable under normal conditions while adaptively evolving during degradation.

Specifically, a compact context representation is first derived from the current node features:

$$\mathbb{r}_t = \mathrm{Avg}\left(\mathbb{Z}_t\right) = \frac{1}{N}\sum\nolimits_{i=1}^{N} \mathbb{Z}_{t,i} \tag{32}$$

where $\mathbb{r}_t \in \mathbb{R}^K$ summarizes the global operating context at time step t. Based on $\mathbb{r}_t$, a context-induced dynamic update is constructed over the prior adjacency. Instead of directly generating a full adjacency matrix from scratch, the proposed method models the graph variation as a low-rank residual perturbation around the prior graph. Specifically, r_t is mapped into two context-dependent low-rank factors:

$$\left[\mathbb{U}_t, \mathbb{V}_t\right] = \psi\left(\mathbb{r}_t\right) \tag{33}$$

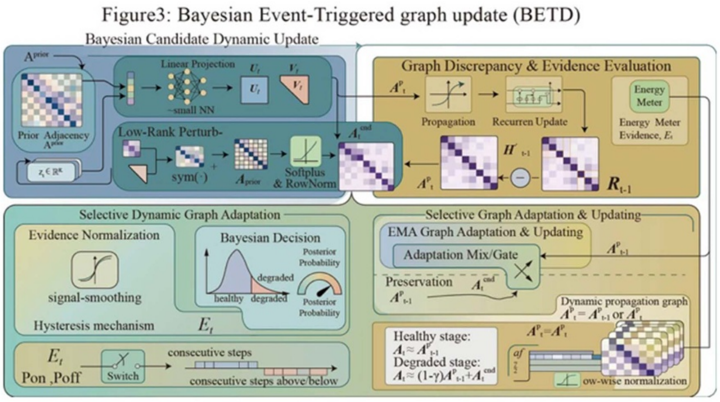


Figure 3. Bayesian event-triggered dynamic graph update mechanism.

where $\psi(\cdot)$ denotes a learnable linear projection, and $\mathbb{U}_t$, $\mathbb{V}_t \in \Re^{N\times q}$, with q being the residual rank. Based on these two factors, the dynamic perturbation is defined as

$$\triangle \mathbb{A}_t = sym(\mathbb{U}_t \mathbb{V}_t^{\mathsf{T}}) \tag{34}$$

where sym$(\cdot)$ denotes the symmetrization operation. This formulation yields a structured low-rank correction that captures degradation-related variations in inter-sensor dependencies. The dynamic update graph is then obtained by applying the residual perturbation to the prior adjacency matrix:

$$\mathbb{S}_t = \mathbb{A}^{prior} + \alpha \cdot \triangle \mathbb{A}_t ,$$

$$\hat{\mathbb{A}}_t = \text{softplus}(\mathbb{S}_t),$$

$$\mathbb{A}_t^{\text{update}} = \mathcal{N}\left(\frac{\hat{\mathbb{A}}_t + \hat{\mathbb{A}}_t^{\mathsf{T}}}{2}\right) \tag{35}$$

where α controls the strength of the dynamic perturbation. softplus$(\cdot)$ ensures non-negativity of edge weights and the final dynamic update adjacency is then obtained by symmetrization, $\mathcal{N}(\cdot)$ denotes row-wise normalization, $\mathbb{A}_t^{\text{update}}$ serves as a candidate graph update, rather than being directly adopted at every time step.

The role of BETD is to determine whether this candidate graph update should be incorporated into the current propagation graph.

From a Bayesian perspective, dynamic graph adaptation can be viewed as a posterior graph change inference problem. Let $\mathbb{A}_t$ denote the latent graph structure at time step t, and let $\mathfrak{D}_{1:t}$ denote the observations accumulated up to time t. Ideally, graph updating should be driven by the discrepancy between the posterior graph distributions at two consecutive time steps, namely $p(\mathbb{A}_t \mid \mathfrak{D}_{1:t})$ and $p(\mathbb{A}_{t-1} \mid \mathfrak{D}_{1:t-1})$.

Accordingly, a generic graph-change trigger may be formulated as

$$\Delta_t = \mathcal{D}(p(\mathbb{A}_t \mid \mathfrak{D}_{1:t}) \parallel p(\mathbb{A}_{t-1} \mid \mathfrak{D}_{1:t-1})) \tag{36}$$

where $\mathcal{D}(\cdot \parallel \cdot)$ denotes a suitable divergence measure between two posterior graph distributions. A large value of Δ_t indicates that the newly inferred graph structure is statistically inconsistent with the graph previously used for propagation, and thus suggests that a graph update should be considered.

In principle, graph change may also be examined at the edge level through posterior edge probabilities. For example, a structural update may be triggered when:

$$\left|\mathcal{P}\left(a_{ij}^{(t)} = 1 \mid \mathfrak{D}_{1:t}\right) - \mathcal{P}\left(a_{ij}^{(t-1)} = 1 \mid \mathfrak{D}_{1:t-1}\right)\right| > \tau_e \tag{37}$$

for some important edge (i,j), where τ_e is a prescribed threshold that is an hyper parameter. This view highlights that dynamic graph updating is fundamentally a Bayesian evidence accumulation problem over graph structure. However, explicitly estimating the full posterior distribution of time-varying graphs and computing the associated posterior divergence at every time step is computationally expensive and statistically unstable in recurrent graph learning. Therefore, instead of directly evaluating Δ_t, the proposed BETD mechanism constructs a tractable surrogate statistic from hidden-state propagation discrepancy, which serves as an online evidence measure for graph inadequacy.

In practie, the graph-propagated hidden state then is computed as:

$$\tilde{\mathbb{H}}_{t-1} = \mathbb{A}_t \mathbb{H}_{t-1}. \tag{38}$$

The subsequent recurrent update of the hidden state follows the same GLSTM formulation introduced in section 2.2. After obtaining the updated hidden state, a graph discrepancy is defined as:

$$\mathbb{R}_t = \mathbb{H}^e{}_t - \tilde{\mathbb{H}}_{t-1} \tag{39}$$

where $\mathbb{H}^e{}_t$ denotes the hidden state after recurrent updating at time step. Based on this residual, the graph-change evidence is measured by the average graph discrepancy energy:

$$\mathfrak{e}_t = \frac{1}{N \cdot D} \left\| \mathbb{R}_t \right\|_F{}^2 \tag{40}$$

where D denotes the hidden dimension. This quantity reflects the extent to which the current hidden dynamics cannot be adequately explained by the graph currently in use.

To normalize the evidence over time, exponential moving averages are maintained for its mean and variance:

$$\mu_t^{\mathfrak{e}} = (1-\lambda)\, \mu_{t-1}^{\mathfrak{e}} + \lambda \mathfrak{e}_t \tag{41}$$

$$\left(\sigma_t^{\mathfrak{e}}\right)^2 = (1-\lambda) \left(\sigma_{t-1}^{\mathfrak{e}}\right)^2 + \lambda \left(\mathfrak{e}_t - \mu_t^{\mathfrak{e}}\right)^2. \tag{42}$$

And the normalized evidence is defined as:

$$\overline{\mathfrak{e}_t} = \frac{\mathfrak{e}_t - \mu_t^{\mathfrak{e}}}{\sigma_t^{\mathfrak{e}}}\ . \tag{43}$$

Graph change detection is then modeled as a Bayesian decision problem. Given a change magnitude parameter $m > 0$, the normalized evidence is converted into a log-likelihood ratio. Assuming that the normalized evidence approximately follows $\mathcal{N}(0,1)$ under the no-change hypothesis and $\mathcal{N}(m,1)$ under the change hypothesis, the corresponding log-likelihood ratio is given by

$$\ell_t = m \overline{\mathfrak{e}_t} - \frac{1}{2} m^2\ . \tag{44}$$

With a prior probability π of graph change, the posterior probability of change is computed as:

$$p_t = \sigma\left(\ell_t + \log \frac{\pi}{1-\pi}\right) \tag{45}$$

where p_t represents the posterior probability that the current graph structure should be updated.

Rather than switching graphs immediately whenever p_t becomes large, a hysteresis mechanism is further adopted to avoid oscillatory updates. Specifically, the model enters the update state only when p_t remains above an upper threshold $\mathrm{D_{on}}$ for n consecutive steps that is a hyper-parameters, and exits the update state only when p_t remains below a lower threshold $\mathrm{D_{off}}$ for S consecutive steps, where $\mathrm{D}_{on} > \mathrm{D_{off}}$.

Once the update state is activated, the propagation graph is progressively adapted toward the candidate dynamic graph by exponential moving averaging:

$$\mathbb{A}_t = \begin{cases} \mathbb{A}_{t-1} \ \textit{no update} \\ (1-\eta)\, \mathbb{A}_{t-1} + \eta\, \mathbb{A}_t{}^{update},\ \textit{updated} \end{cases} \tag{46}$$

where $\eta \in (0,1]$ is the graph adaptation rate. After each update, $\mathbb{A}_t$ is row-normalized to preserve a valid propagation structure.

Therefore, unlike conventional dynamic graph methods that update the adjacency matrix at every time step, the proposed BETD mechanism performs selective, evidence-driven graph adaptation. This design improves structural robustness while retaining the ability to capture degradation-induced evolution of inter-sensor dependencies.

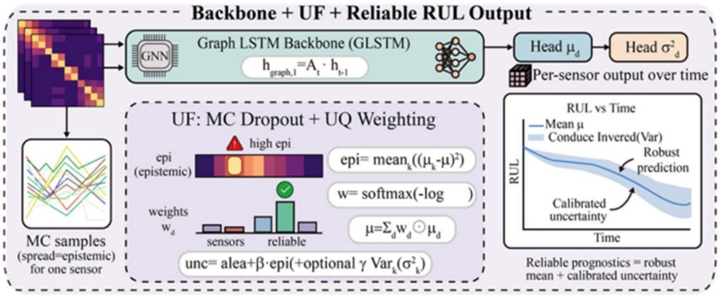


Figure 4. Diagram of uncertainty-aware dynamic fusion module.

3.4. UDF module

In multi-sensor prognostics, the reliability of different sensors is often heterogeneous and time-varying due to measurement noise, operational disturbances, and localized degradation. As a result, simply averaging sensor-level predictions may lead to suboptimal prognostic performance, especially when unreliable sensors dominate the fusion outcome. To address this issue, an UDF module is proposed, which adaptively aggregates sensor-level predictions according to their predictive reliability estimated from uncertainty quantification. Specifically, the proposed UDF exploits MC dropout to estimate both aleatoric and epistemic uncertainties for each sensor, constructs a comprehensive uncertainty score, and then assigns larger fusion weights to sensors with higher predictive precision.

Figure 4 illustrates the overall architecture of the proposed UDF module, its corresponding pseudocode is in the appendices.

Assume that the monitored system is equipped with N sensors. For the nth sensor at time step t, K stochastic forward passes are performed under MC dropout, yielding a set of predictive means and predictive variances, denoted by $\mu_{t,n}^{(k)}$ and $\sigma_{t,n}^{2(k)}$, respectively, where $k = 1, 2, \ldots, K$. Based on these stochastic predictions, the sensor-level predictive mean is defined as: $\bar{\mu}_{t,n} = \frac{1}{K}\sum_{k=1}^{K} \mu_{t,n}^{(k)}$. The aleatoric uncertainty, which characterizes the intrinsic noise associated with the sensor observation, is estimated by averaging the predictive variances over all stochastic passes: $\mu_{t,n}^{\text{ale}} = \frac{1}{K}\sum_{k=1}^{K} \sigma_{t,n}^{2(k)}$.

In addition, the epistemic uncertainty, which reflects the model uncertainty arising from limited knowledge of the degradation process, is quantified by the dispersion of the predictive means across MC dropout samples: $\mu_{t,n}^{\text{epi}} = \frac{1}{K}\sum_{k=1}^{K} \left(\mu_{t,n}^{(k)} \bar{\mu}_{t,n}\right)^2$.

To jointly account for data noise and model uncertainty, a comprehensive uncertainty measure is constructed by equation (20) as: $\mu_{t,n} = \mu_{t,n}^{\text{ale}} + \beta\, \mu_{t,n}^{\text{epi}}$, where β is a non-negative balancing coefficient that controls the contribution of epistemic uncertainty. This formulation implies that the predictive reliability of a sensor is jointly determined by the inherent stochasticity in the data and the confidence of the model with respect to that sensor.

On this basis, a precision score is assigned to each sensor as the inverse of its comprehensive uncertainty: $\mathrm{D}_{t,d} = \frac{1}{\mu_{t,n} + \varepsilon}$, where ε is a small positive constant introduced to ensure numerical stability. The dynamic fusion weight of the dth sensor is then obtained by normalizing the precision scores across all sensors: $\omega_{t,n} = \frac{\mathrm{D}_{t,n}}{\sum_{j=1}^{D} \mathrm{D}_{t,j}}$.

Accordingly, sensors associated with lower comprehensive uncertainty are assigned larger fusion weights, thereby contributing more strongly to the final prognostic output.

Using the resulting weights, the fused predictive mean at time step t is formulated as: $\mu_t^{\text{UDF}} = \sum_{n=1}^{N} \varpi_{t,n} \bar{\mu}_{t,n}$.

To obtain a rigorous uncertainty estimate for the fused prediction, the predictive variance is derived through second-order moment propagation. Specifically, the fused predictive variance is expressed as:

$$\left(\sigma_t^2\right)^{\text{UDF}} = \sum_{n=1}^{N} \varpi_{t,n} \left(u_{t,n} + \bar{\mu}_{t,n}^2\right) - \left(\mu_t^{\text{UDF}}\right)^2. \tag{47}$$

The above provides a more complete characterization of the fused uncertainty than a simple weighted average of sensor-level variances.

In particular, it captures not only the uncertainty associated with each individual sensor prediction, but also the disagreement among sensor-wise predictive means. Through this mechanism, the proposed UDF module adaptively emphasizes sensor channels with higher predictive reliability and suppresses those with elevated uncertainty, thereby improving the robustness and credibility of the final prognostic result.

3.5. Evaluation metrics

The evaluation metrics adopted in this paper were selected with reference to a recent comprehensive review on machine RUL prediction, in which quantitative metrics are categorized from the perspectives of accuracy, prediction confidence, and prediction stability (Zhou *et al* 2024). Following this rationale, three representative metrics are employed in this study, namely root mean square error (RMSE, see equation (48)), NASA Score (see equation (49)), and Mean absolute percentage error (MAPE, see equation (50)):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \bar{y}_i\right)^2} \tag{48}$$

$$SCORE = \begin{cases} \sum\limits_{i=1}^{N}\left(e^{-\frac{\bar{y}_i - y_i}{13}} - 1\right), \bar{y}_i < y_i \\ \sum\limits_{i=1}^{N}\left(e^{\frac{\bar{y}_i - y_i}{10}} - 1\right), \bar{y}_i \geqslant y_i \end{cases} \tag{49}$$

$$MAPE = \frac{1}{EOL - FPT}\sum_{i=FPT}^{EOL}\left|\frac{\bar{y}_i - y_i}{y_i}\right| \times 100\%. \tag{50}$$

Where EOL denotes the end-of-life point and FPT denotes the first prediction time. N denotes the total number of test samples, and y_i and $\bar{y}_i$ represent the model's predicted output and the corresponding ground-truth value, respectively.

4. Case study

4.1. Dataset and experimental scenarios

The NASA C-MAPSS dataset, that is widely used for RUL prediction validation, is utilized to construct multiple testing scenarios for evaluating the proposed model's prognostic capability. It comprises four sub-datasets (FD001–FD004), each representing distinct combinations of operating conditions and fault modes. Every sub-dataset is divided into training and test sets, as summarized in table 1. In this study, 14 sensors are selected, consistent with the sensor configuration used in (Liu *et al* 2022).

The proposed model is evaluated under three settings. Scenario I corresponds to the standard condition, where RUL is predicted on the original C-MAPSS test set without any perturbation. Scenario II examines robustness to sensor faults. In this setting, one sensor channel in each test sequence is intentionally corrupted using a predefined fault pattern; in this study, a stuck-at fault is simulated on Sensor 1. Specifically, after a certain time point, the readings of Sensor 1 are fixed to a constant value to mimic a sensor failure. By simulating this realistic single-sensor fault, the model is further assessed for its ability to provide reliable predictions under abnormal operating conditions.

To improve reproducibility, the key implementation settings are briefly summarized here, including the data split, input sequence length, and optimization setup. More detailed training configurations, pseudocode, and model hyperparameter settings are provided in appendix A–C. Training efficiency is reflected by the number of model parameters (In appendix C), which is independent of the computing device. In addition, a sensitivity analysis of the key hyperparameters, including the event-triggering threshold, the update rate, and the fusion balance coefficient, is provided in appendix C to examine their influence on the prediction performance across FD conditions. For testing, the corresponding inference-time statistics are reported in appendix D.

4.2. Comparison with several state-of-the-art (SOTA) prognostic models

First, the proposed model is compared with SOTA prognostic approaches, including KLN (Zhou *et al* 2026), DMAN (Qian *et al* 2026), AttnPINN (Liao *et al* 2023), STA-HPINN (Jiang *et al* 2025), TCN-AE-Wiener (Guo *et al* 2026), EWM- BiLSTM (Qiu *et al* 2025), LSTM-E (Nemani *et al* 2022), SAG-CAPsNet (Li *et al* 2024), STGNN (Luo *et al* 2024), STAGNN (Luo *et al* 2024), HAGCN (Zhang *et al* 2020) and ASTGCNN (Zhang *et al* 2020). These representative methods cover both physics-informed prognostics and graph-based spatiotemporal modeling frameworks.

Table 1. C-MAPSS FD001–FD004 Dataset Overview.

C-MAPSS	FD001	FD002	FD003	FD004
Training	100	260	100	249
Testing	100	259	100	248
Conditions	1	6	1	6
Fault modes	1	1	2	2
Training size	362	378	525	543
Testing size	100	259	100	248

Table 2 compares the proposed framework with several representative state-of-the-art prognostic models on FD001–FD004. Since RMSE, Score, and MAPE are all lower-is-better metrics, the superiority of a model can be evaluated by the extent to which it reduces these three indicators relative to competing methods. From this perspective, the proposed framework demonstrates the most consistent overall advantage across datasets with different levels of degradation complexity.

For RMSE, the proposed method achieves the best performance on all four datasets, indicating that it provides the most accurate point prediction overall. In the relatively simple FD001 scenario, the proposed model reduces RMSE by about 9.9% compared with AttnPINN, by 15.6% compared with ASTGCNN, and by more than 34% compared with the conventional LSTM-based baseline, showing a clear advantage even under relatively clean degradation conditions. In FD002, where multiple operating conditions increase distributional variability, the improvement margin becomes smaller, but the proposed framework still maintains the lowest RMSE, with a reduction of about 2.1% relative to AttnPINN. This suggests that although competing attention-based methods are already effective in handling operating-condition shifts, the proposed framework still offers a more stable representation. In FD003, the RMSE advantage becomes more pronounced again: compared with AttnPINN and ASTGCNN, the proposed method reduces RMSE by about 11.1% and 28.1%, respectively, indicating a stronger ability to capture nonlinear degradation patterns under multiple fault modes. In the most complex FD004 dataset, the proposed method still ranks first in RMSE, although the margin over the strongest competitors is relatively moderate, which indicates that all advanced models face increasing difficulty under coupled operating and fault variations, while the proposed framework remains the most accurate overall.

For the Score metric, the proposed framework also achieves the lowest value on all four datasets, which further confirms that its advantage is not limited to reducing average prediction error, but also extends to improving degradation timing estimation. In FD001, the proposed model lowers the Score by about 7.4% compared with AttnPINN and by 31.9% compared with ASTGCNN. In FD002, it still reduces the Score by about 8.1% relative to AttnPINN, showing that the model remains robust even when operational regimes become more diverse. The most substantial gain appears in FD003, where the proposed method reduces the Score by about 18.3% compared with AttnPINN and by more than 84% compared with ASTGCNN. This large margin indicates a much stronger capability in identifying critical degradation transitions and avoiding severe prediction bias in complex fault-mode scenarios. In FD004, the proposed model continues to outperform all compared approaches, achieving about an 8.0% reduction over AttnPINN and a much larger reduction over more conventional deep learning baselines. Therefore, from the perspective of Score, the proposed method shows the strongest and most stable advantage across all degradation scenarios.

For MAPE, the proposed framework also performs competitively and achieves the lowest value on three out of four datasets, demonstrating strong relative-error control. In FD001, the proposed method reduces MAPE by about 14.4% compared with AttnPINN, by 18.7% compared with ASTGCNN, and by over 43% relative to the LSTM baseline, indicating a substantial improvement in percentage-based prediction accuracy. In FD002, the improvement over the strongest competitor is smaller, with about 1.4% reduction relative to AttnPINN, but the proposed framework still remains the best-performing model. In FD003, the proposed model does not achieve the lowest MAPE; instead, it is slightly higher than the best competitor by about 2.1%. However, this small disadvantage in MAPE should be interpreted together with the fact that the proposed model still achieves the best RMSE and by far the best Score on this dataset, suggesting that its overall prognostic performance remains superior, especially in terms of absolute accuracy and degradation trend tracking. In FD004, the proposed method again achieves the lowest MAPE, with a reduction of about 2.8% relative to AttnPINN and nearly 19% relative to ASTGCNN, confirming its strong robustness under the most challenging conditions.

Table 2. Comparison with the SOTA prognostic models in scenario. I

	RMSE				Score				MAPE			
Approach	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
AttnPINN	13.08	11.32	11.07	12.76	258	654	257	798	15.24	14.37	14.25	18.79
STA-HPINN	13.47	11.58	11.36	13.12	287	701	279	846	15.88	14.69	**13.50**	18.60
TCN-AE-Wiener	17.62	15.84	16.23	16.98	498	1182	1218	1698	23.21	25.37	23.85	29.78
EWM- BiLSTM	18.14	16.37	17.05	17.82	542	1248	1305	1796	23.21	26.93	27.03	31.59
LSTM-E	17.89	15.97	16.82	17.41	521	1203	1247	1749	22.92	26.96	25.89	31.17
SAG-CAPsNet	14.18	11.96	12.24	13.58	326	748	301	902	18.35	16.48	14.44	21.25
STGNN	15.52	12.54	15.39	15.11	470	845	1534	1375	19.96	20.23	25.80	24.96
STAGNN	14.81	12.21	14.87	14.72	418	782	1446	1319	19.89	18.32	24.48	24.64
HAGCN	14.32	11.88	14.12	14.26	382	721	1378	1256	17.48	17.21	23.71	22.85
ASTGCNN	13.97	11.63	13.68	13.91	351	689	1321	1208	16.05	14.44	22.09	22.48
KLN	12.68	11.21	10.83	13.09	255	637	316	883	14.63	16.37	14.92	20.20
DMAN	13.33	11.67	11.26	12.97	310	660	311	838	14.98	15.34	16.12	19.32
Proposed	**11.79**	**11.08**	**9.84**	**12.63**	**239**	**601**	**210**	**734**	**13.05**	**14.17**	13.79	**18.26**

These results indicate that the proposed framework does not merely improve prediction accuracy under specific conditions. Instead, it provides consistent gains across increasing levels of degradation complexity, demonstrating superior adaptability, structural modeling capability, and uncertainty calibration.

Table 3 further shows that the proposed framework maintains strong robustness when sensor measurements are corrupted. Although most competing methods exhibit performance degradation under this setting, the proposed model still preserves its overall advantage across the four datasets. In terms of RMSE, it continues to achieve the best performance on FD001, FD003, and FD004, while remaining highly competitive on FD002, indicating that its point prediction accuracy is less affected by abnormal sensor disturbances. For the Score metric, the proposed model still ranks first on all four datasets, suggesting that its superiority in degradation trend tracking and failure timing estimation remains stable even under attacked inputs. In addition, the proposed framework achieves the best MAPE on FD002, FD003, and FD004, and stays very close to the best-performing competitor on FD001, which demonstrates stronger control of relative prediction error under corrupted sensing conditions. These results indicate that the proposed method not only performs well under normal conditions, but also retains reliable prognostic capability when the input signals become disturbed, highlighting its robustness and adaptability in more challenging environments.

4.3. Uncertainty analysis for decision-making fusion

Epistemic uncertainty reflects the reliability of each sensor's RUL prediction, which motivates the design of UDF for decision-level information fusion. In this section, the proposed method investigates the epistemic uncertainty associated with the RUL predictions from individual sensors and illustrates how it evolves along the pseudo-degradation process under different degradation conditions (FD001–FD004).

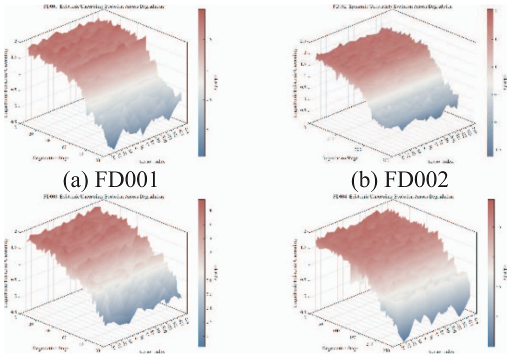
As shown in figure 5, a broadly consistent trend can be observed across FD001 to FD004. The epistemic uncertainty generally decreases as the degradation process progresses. In the early stage, when the system remains close to a healthy condition, the uncertainty level is relatively high for most sensors. This suggests that the model is less confident when degradation-related features are still weak, diffuse, or not yet clearly distinguishable from normal operational variability. At this stage, the separability between different health conditions is limited, which can naturally result in higher epistemic uncertainty. As degradation advances, latent failure-related patterns become more pronounced and structured, making them easier for the model to identify and learn. Consequently, the uncertainty surface gradually slopes downward, particularly in the mid-to-late degradation stages, where the model exhibits increased confidence in characterizing the system state. Although the overall trend is similar across all four datasets, some differences can still be observed. FD002 and FD004 show slightly larger variability across sensors than FD001 and FD003, which may be associated with their more complex operating conditions. Nevertheless, the general decreasing tendency remains visible in all cases. It should be noted, however, that this behavior should not be interpreted as a universal monotonic rule for all possible degradation or failure modes. In some scenarios, especially during fault-transition stages or when multiple degradation mechanisms coexist, epistemic uncertainty may increase locally due to temporarily more complex and less separable system responses. Therefore, the pattern observed should be understood as a consistent empirical trend in the studied datasets, indicating that epistemic uncertainty is meaningfully related to degradation progression and can serve as an informative measure of model confidence.

The relationship between the epistemic uncertainty and the prediction error of each sensor is also examined.

Figure 6 presents the Pearson correlation coefficients between epistemic uncertainty and prediction error for all sensors under four operating conditions including FD001–FD004. For FD001–FD003, consistently positive and statistically significant correlations are observed across all sensors. The correlation coefficients generally fall within the range of approximately 0.25–0.45, indicating a moderate association between uncertainty and error. This suggests that higher epistemic uncertainty tends to coincide with larger prediction errors, reflecting an alignment between the model's confidence estimation and its actual predictive performance. Although the correlation strength decreases under the more complex FD004 condition, the relationship remains positive across sensors. This reduction is reasonable given the increased operational variability in FD004, where additional sources of uncertainty may weaken the direct association between epistemic uncertainty and prediction error. Nevertheless, the consistent positive trend across datasets indicates that epistemic uncertainty is not randomly fluctuating but systematically related to prediction reliability. Overall, these findings indicate that epistemic uncertainty carries meaningful information about prediction reliability and can reasonably serve as a proxy for prediction error. This provides empirical support for using uncertainty as an adaptive weighting factor in the proposed UDF framework.

Table 3. Comparison with the SOTA prognostic models in scenario II.

	RMSE				Score				MAPE			
Approach	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
AttnPINN	13.19	11.41	11.22	12.92	271	751	255	816	16.37	14.91	15.56	18.99
STA-HPINN	13.68	11.74	11.58	13.35	309	792	279	874	16.92	14.79	15.48	18.15
TCN-AE-Wiener	18.54	16.72	17.18	17.94	575	1340	1365	1848	23.08	26.24	27.21	30.72
EWM- BiLSTM	19.33	17.42	18.06	18.83	620	1415	1448	1952	25.13	28.35	28.02	32.35
LSTM-E	18.67	16.58	17.74	18.21	593	1357	1392	1889	24.24	27.78	27.39	31.63
SAG-CAPsNet	14.44	12.26	12.53	13.91	348	828	321	946	16.59	15.96	16.84	19.52
STGNN	16.23	13.01	16.04	15.78	505	973	1541	1358	18.65	17.43	25.09	25.18
STAGNN	15.52	12.66	15.43	15.21	460	902	1463	1404	17.92	17.22	24.51	24.46
HAGCN	15.01	12.32	14.71	14.83	420	846	1394	1326	17.55	16.16	24.33	24.34
ASTGCNN	14.69	12.09	14.32	14.55	398	812	1341	1284	16.74	15.34	23.69	22.69
KLN	13.36	11.62	11.17	13.22	296	667	268	834	16.89	15.23	15.32	21.17
DMAN	13.58	**11.35**	11.64	13.04	280	674	301	823	**16.01**	16.59	16.51	20.69
Proposed	**11.95**	11.77	**10.14**	**12.84**	**270**	**664**	**218**	**799**	16.22	**14.69**	**14.41**	**18.67**



(a) FD001 (b) FD002

(c) FD003 (d) FD004

Figure 5. Epistemic uncertainty evolution across degradation.

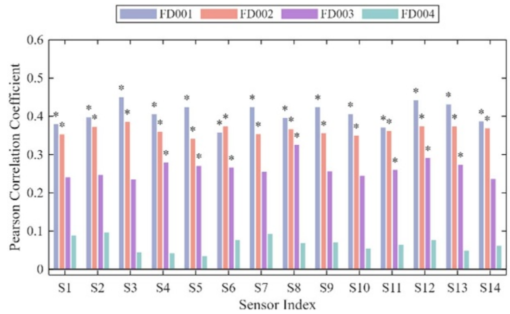


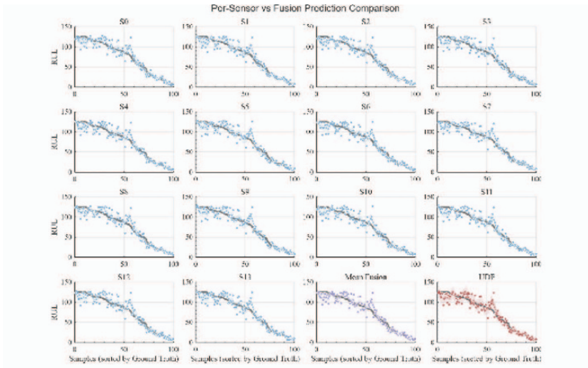
Figure 6. Pearson correlation coefficients between epistemic uncertainty and prediction error.

The distribution of the RUL predictions associated with each sensor at each degradation point is also visualized in figure 7.

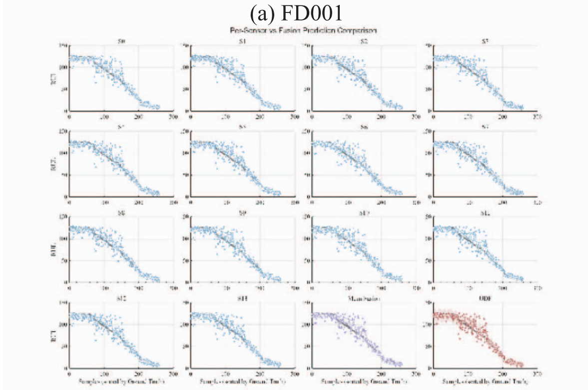
Figure 7 across the 14 individual sensors reveals substantial inter-sensor variability, particularly in the mid-to-late degradation regime. Several sensors exhibit systematic bias and stage-dependent fluctuations, indicating that single-sensor inference may lead to unstable and potentially over-confident predictions. Mean Fusion partially mitigates this variability by averaging sensor outputs, resulting in smoother trajectories and reduced oscillation. However, it treats all sensors equally and does not account for their reliability differences, leading to residual bias in critical degradation stages. In contrast, UDF demonstrates the most consistent alignment with ground truth across all stages. By incorporating epistemic uncertainty into the fusion process, UDF effectively suppresses unreliable sensor contributions while amplifying informative ones. This adaptive weighting mechanism significantly enhances prediction stability and reliability calibration, particularly near failure. These observations confirm that fusion is essential, and that UDF outperforms naive averaging strategies.

While the previous experiments demonstrate that UDF improves prediction stability under normal operating conditions, real-world monitoring systems are often subject to sensor malfunction, noise corruption, or even malicious data injection. In such scenarios, individual sensor measurements may become unreliable or adversarially biased. To evaluate the robustness of the fusion strategies, sensor attack scenarios are further simulated, where selected sensors are intentionally corrupted. Figure 8 aims to assess whether the fusion framework can effectively suppress abnormal sensor contributions and maintain reliable RUL estimation.

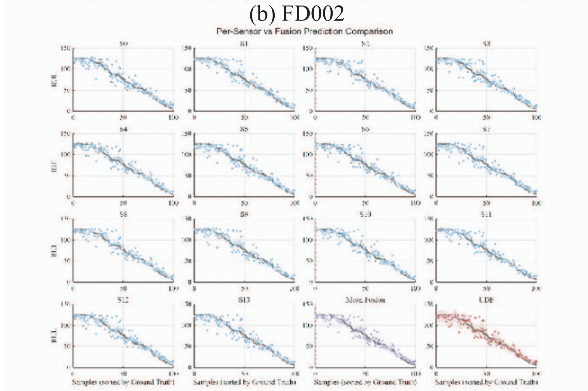
Figure 8 demonstrates the robustness comparison between Mean Fusion and UDF under sensor attack conditions. When selected sensors are corrupted, Mean Fusion exhibits noticeable deviation from ground truth, particularly in the mid-to-late degradation regime. Since Mean Fusion equally aggregates all sensor predictions, abnormal sensor outputs directly bias the fused result, resulting in amplified fluctuation and stage-dependent prediction drift. In contrast, UDF maintains significantly improved stability and alignment with ground truth across all degradation stages. The uncertainty-aware weighting mechanism effectively suppresses unreliable sensor contributions, thereby mitigating the impact of corrupted inputs. Moreover, UDF provides more consistent uncertainty calibration, as reflected by its confidence



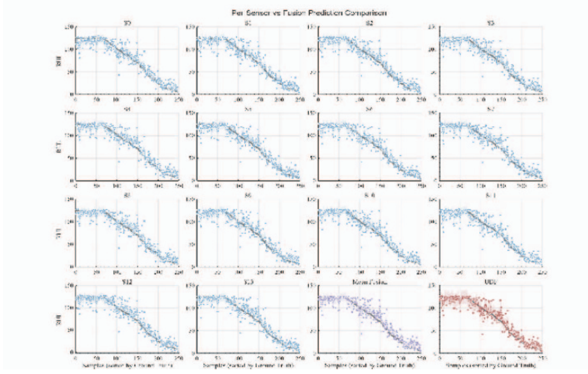
(a) FD001



(b) FD002



(c) FD003



(d) FD004

Figure 7. Prognostic performance under scenario I.

interval behavior under attack conditions. These results confirm that reliability-aware fusion is essential for practical deployment, where sensor corruption and abnormal measurements are unavoidable.

4.4. DA analysis for performance improvement

In this section, sensor adjacency is visualized to analyze the effect of the proposed BETD module during the long-term degradation process. Figures 9 and 10 compare the sensor graph structures under different time steps on FD001 and FD002, respectively.

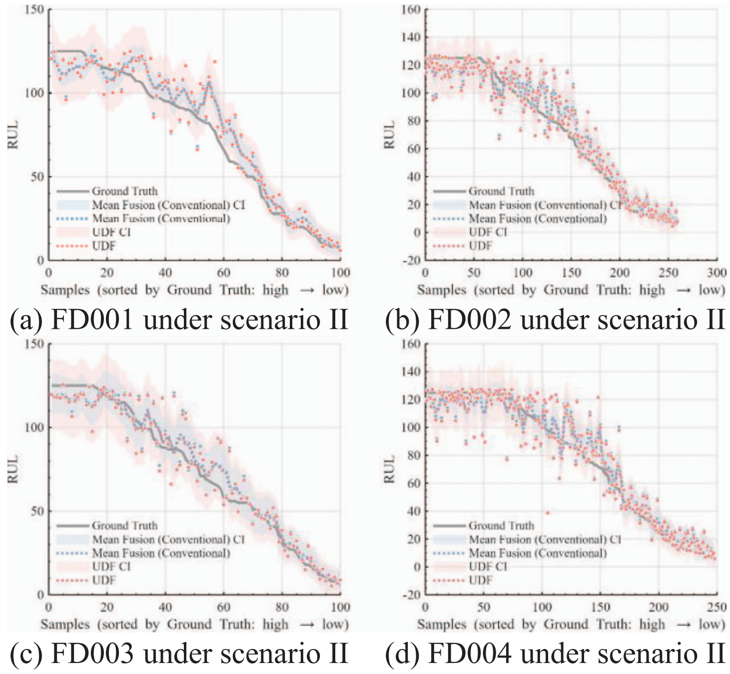


Figure 8. Prognostic performance under scenario II.

As shown in figure 9, the impact of the proposed BETD mechanism on graph evolution can be clearly observed by comparing the adjacency patterns under the same degradation scenario. Without BETD, the learned adjacency matrices exhibit widespread and frequent fluctuations across samples, as shown in figure 9(a). This behavior is further confirmed by the max–min variation (figure 9(b)) and standard deviation map (figure 9(c)), where a large portion of sensor pairs presents noticeable variations. Such globally distributed changes indicate that the graph structure is updated in a fully data-driven manner at every time step, making it sensitive to instantaneous observation noise and short-term fluctuations. As a result, the learned inter-sensor dependencies are less structured and may not accurately reflect the gradual nature of degradation. In contrast, after introducing BETD, the graph evolution becomes significantly more structured and selective. As shown in figure 9(d), the sample-wise adjacency matrices exhibit higher consistency over time, with only limited regions undergoing visible changes. This is consistent with the design of BETD, where graph updates are triggered only when sufficient evidence of structural inconsistency is accumulated. Correspondingly, the max–min variation (figure 9(e)) is concentrated on a smaller subset of sensor pairs, while most connections remain stable. Meanwhile, the standard deviation map (figure 9(f)) shows a more localized and organized pattern, indicating smoother and more controlled graph evolution.

This difference can be directly attributed to the Bayesian evidence-driven update mechanism in BETD. By measuring the discrepancy between hidden state propagation and actual system dynamics, BETD avoids unnecessary updates and suppresses noise-induced structural oscillations. The hysteresis strategy further prevents frequent switching between graph states, ensuring temporal consistency.

Figure 10 presents the effect of BETD on graph evolution under the FD002 scenario, which involves multiple operating conditions and stronger non-stationarity, can be clearly observed. Compared with FD001, figure 10 shows that the advantage of BETD becomes more evident under the FD002 scenario, where multiple operating conditions introduce stronger non-stationarity. Without BETD, the graph variations become more irregular and scattered, indicating that the learned dependencies are more easily disturbed by condition-induced fluctuations. After incorporating BETD, the adjacency evolution remains comparatively more organized, and the edge variations are still confined to a relatively limited subset of sensor pairs. This suggests that BETD is able to preserve the selectivity of graph adaptation even in the presence of operating-condition changes, thereby helping the model distinguish true degradation-related dependency evolution from condition-induced variability.

Figures 11 and 12 illustrate the model performance under attack scenarios.

Figure 11 further investigates the behavior of the learned sensor graph on FD003 under noise attack scenarios. Without BETD, the learned adjacency matrices exhibit highly irregular and unstable patterns across samples, as illustrated in figure 11(a). Compared with the previous scenarios, the graph structure

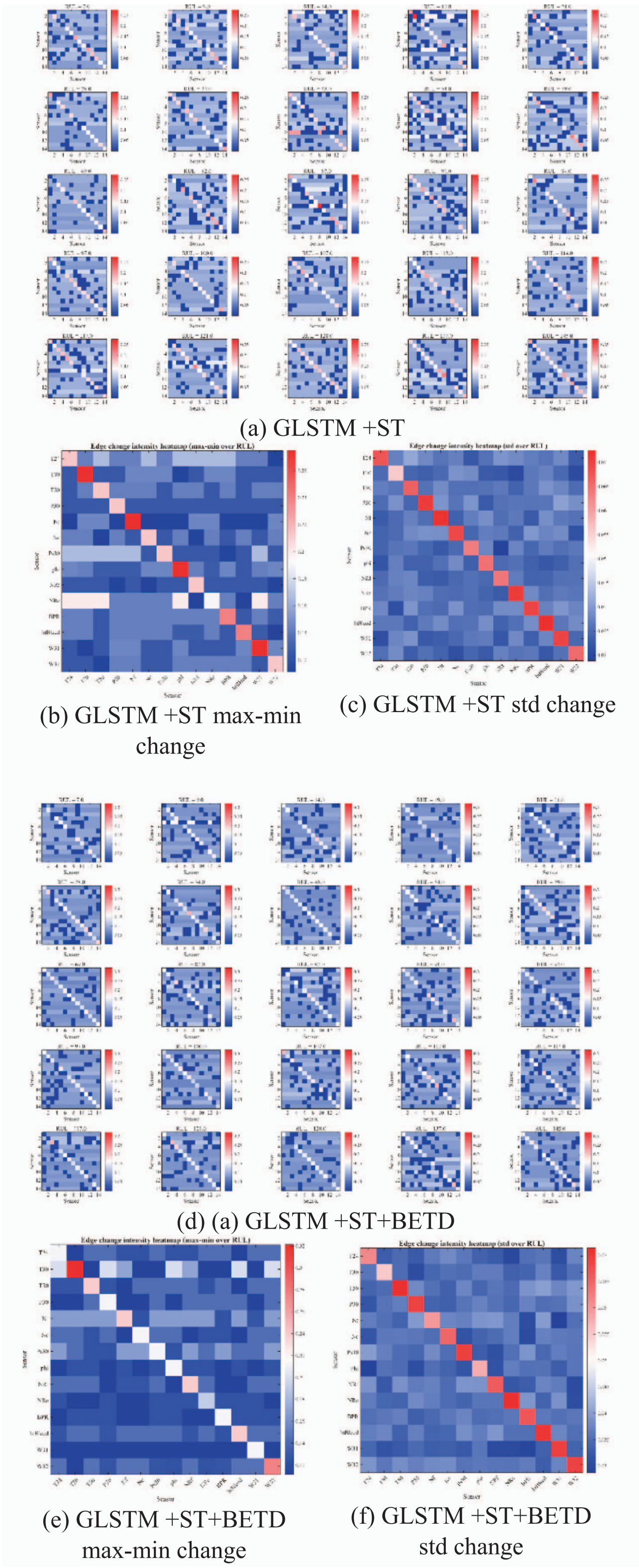
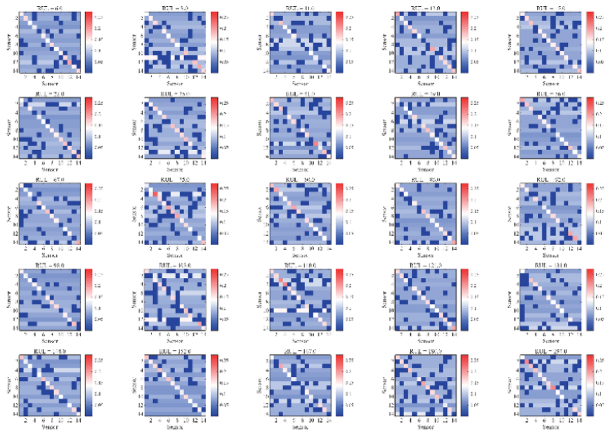
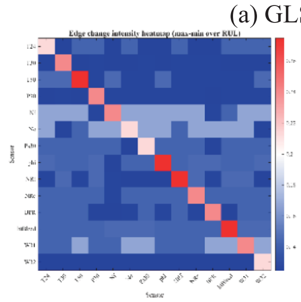


Figure 9. FD001 under clean scenario.

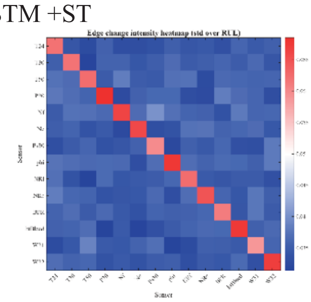
shows more abrupt and scattered variations, which are not aligned with the expected gradual degradation process. This is further reflected in the max–min variation (figure 11(b)) and standard deviation map (figure 11(c)), where strong fluctuations are distributed across a wide range of sensor pairs. Such behavior suggests that the graph learning mechanism is heavily influenced by corrupted or abnormal sensor readings, leading to spurious dependency modeling and structural instability.In contrast, after introducing BETD, the graph evolution becomes significantly more robust and structured despite the



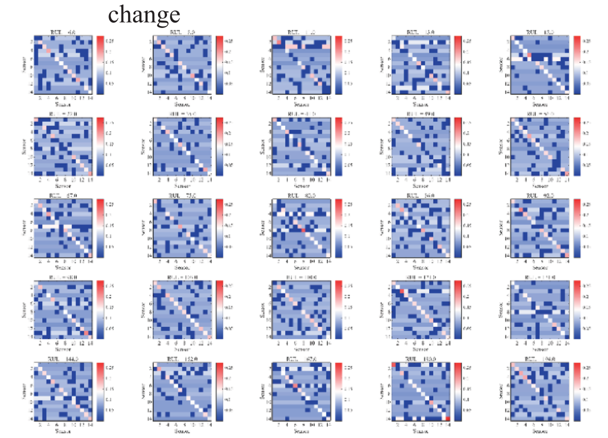
(a) GLSTM +ST



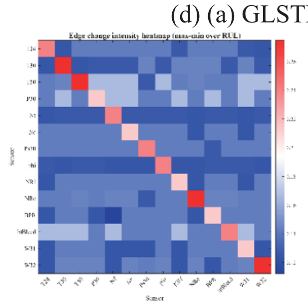
(b) GLSTM +ST max-min change



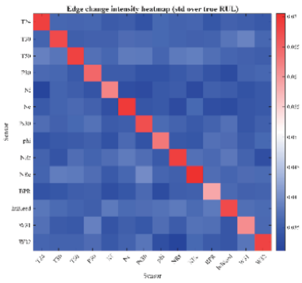
(c) GLSTM +ST std change



(d) (a) GLSTM +ST+BETD



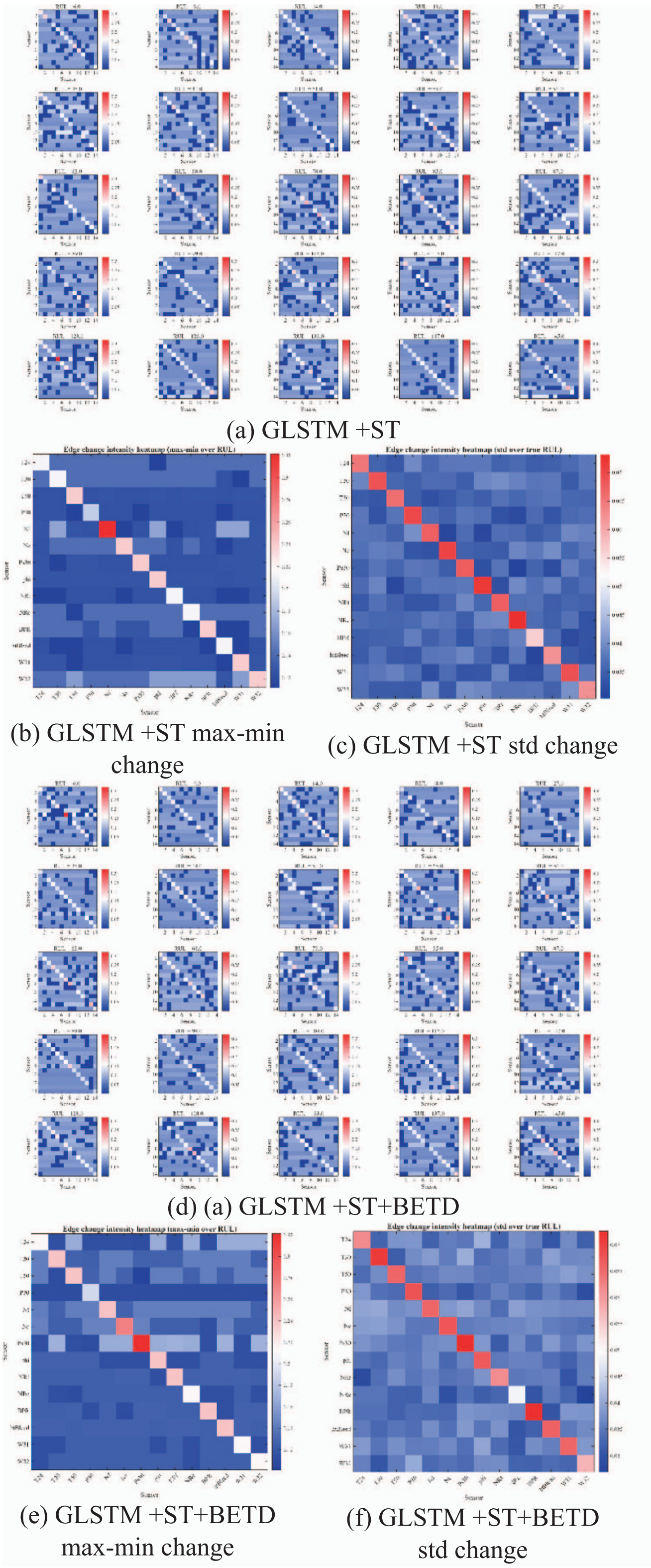
(e) GLSTM +ST+BETD max-min change



(f) GLSTM +ST+BETD std change

Figure 10. FD002 under clean scenario.

presence of attacks. As shown in figure 11(d), the sample-wise adjacency matrices exhibit improved consistency, with fewer abrupt and noisy variations. The max–min variation (figure 11(e)) becomes more concentrated, indicating that only a limited subset of sensor relationships is affected. Meanwhile, the standard deviation map (figure 11(f)) shows a more organized and localized pattern, suggesting that the influence of attacked sensors is effectively constrained rather than propagating across the entire graph. This improvement can be attributed to the Bayesian evidence-driven update mechanism in BETD. Since



(a) GLSTM +ST

(b) GLSTM +ST max-min change

(c) GLSTM +ST std change

(d) (a) GLSTM +ST+BETD

(e) GLSTM +ST+BETD max-min change

(f) GLSTM +ST+BETD std change

Figure 11. FD003 under attack scenario.

graph updates are triggered only when the hidden-state discrepancy provides consistent and statistically significant evidence, transient or abnormal perturbations introduced by attacks are less likely to activate unnecessary graph changes. Furthermore, the hysteresis strategy prevents rapid switching of graph structures under unstable conditions, thereby enhancing robustness.

This robustness is further confirmed in figure 12 under the FD004 scenario, which combines sensor attacks with multiple operating conditions and is therefore more challenging than FD003. Without

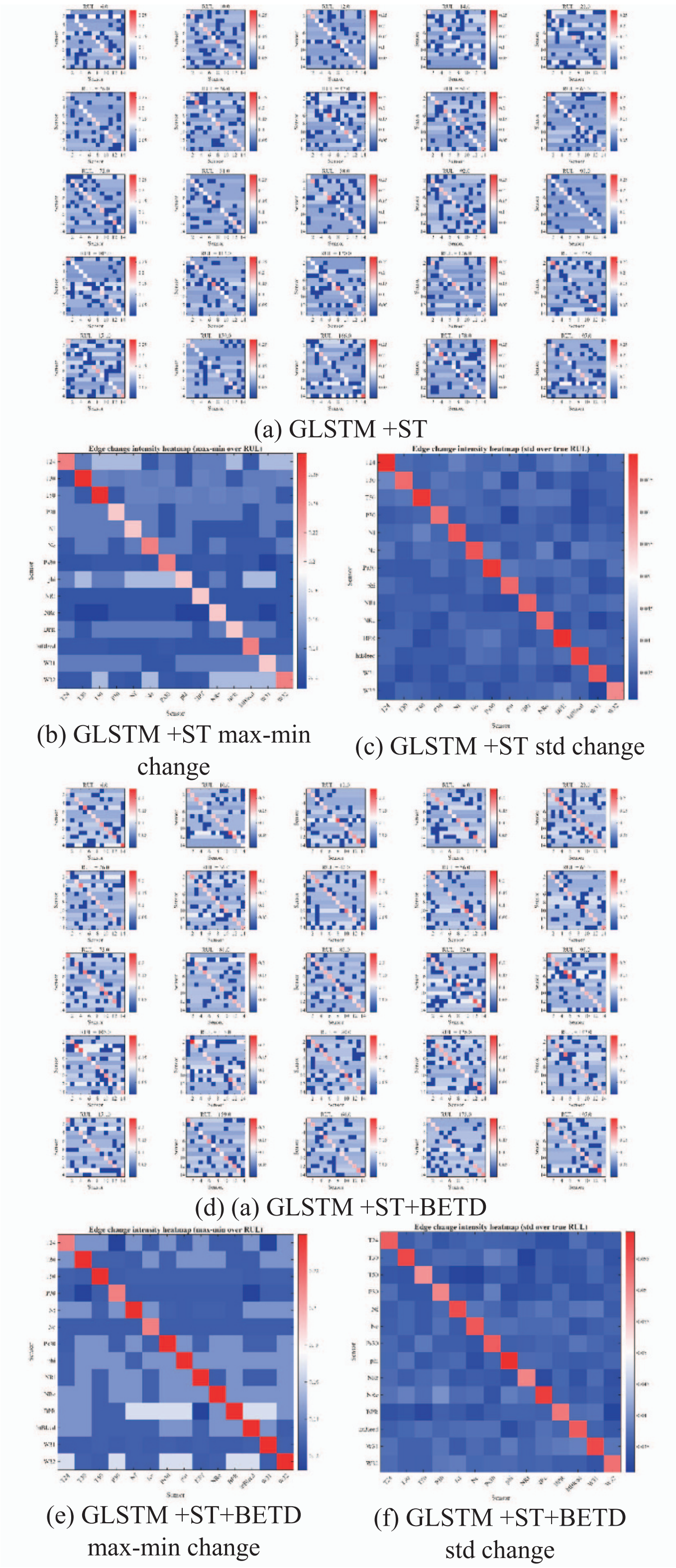


Figure 12. FD004 under attack scenario.

BETD, the graph variations become even more chaotic, reflecting the joint interference of condition shifts and attack perturbations. Nevertheless, with BETD, the adjacency evolution still remains relatively structured, and the edge variation statistics continue to show a more concentrated pattern. This indicates that BETD not only mitigates attack-induced disturbances, but also prevents their interaction with operating-condition variability from causing excessive graph instability. As a result, the learned graph remains more physically meaningful and better aligned with the underlying degradation process even in the most complex scenario.

4.5. Ablation study

This section aims to systematically investigate the individual contributions of the key components in the proposed framework through a series of ablation experiments. To ensure a fair comparison, GLSTM is adopted as the baseline model, since it is a widely used architecture for modeling ST dependencies in sensor-driven prognostics tasks. By progressively enabling or disabling different modules, the ablation study reveals how each component contributes to the overall prognostic performance of the proposed method.

Table 4 presents the ablation results under Scenario I. It can be observed that the ST module provides the most direct contribution to prediction accuracy, as its introduction leads to substantial reductions in both RMSE and MAPE across all datasets compared with the baseline GLSTM. This indicates that enhanced ST representation learning is essential for capturing degradation evolution more accurately. The BETD module further improves the prognostic performance, particularly in terms of the Score metric, with more evident gains on FD004, where the operating conditions and degradation patterns are relatively more complex. This suggests that dynamic topology modeling helps better characterize structural dependencies among sensor channels. The UDF module, when introduced individually, brings relatively moderate improvements in RMSE, but it still contributes to lower MAPE and better Score on several datasets, indicating its role in refining feature fusion and improving prediction quality. When all three modules are integrated, the complete model achieves the best overall results on most datasets, demonstrating that the proposed framework benefits from the complementary effects of ST modeling, structural dependency learning, and UDF.

These results indicate that accuracy enhancement and uncertainty calibration are fundamentally different objectives, and their joint optimization through modular design leads to superior prognostic performance. Table 5 give the prognostic performance under scenario II.

Table 5 reports the ablation results under Scenario II, where sensor attack conditions make the prognostic task more challenging. Compared with Scenario I, the baseline GLSTM suffers from clear performance degradation in RMSE, Score, and MAPE, indicating its sensitivity to corrupted sensor inputs. Under this condition, the ST module still remains the dominant contributor to accuracy improvement, consistently reducing both RMSE and MAPE across datasets and showing strong robustness in disturbed environments. The BETD module provides additional gains, especially in the Score metric, suggesting that explicit modeling of dynamic sensor relationships enhances the structural resilience of the model under attack. The UDF module further improves performance by refining feature fusion under uncertain and perturbed inputs. Although its standalone contribution is still less pronounced than that of ST, its effect becomes more evident when combined with ST and BETD, leading to more stable and accurate predictions. The full model again achieves the best overall performance, indicating that the three modules remain complementary even under adverse sensor conditions.

5. Conclusions and future works

5.1. Conclusions

This study presented an uncertainty-aware prognostic framework for RUL prediction from multivariate monitoring data. The framework was developed to improve the modeling of temporal degradation patterns, evolving inter-sensor dependencies, and sensor-level predictive uncertainty within a unified prediction pipeline.

The main findings of this study can be summarized as follows:

1. A spatial–temporal modeling architecture was developed for multivariate RUL prediction. By combining multi-scale temporal feature extraction with graph-based recurrent modeling, the proposed architecture was able to capture both temporal degradation information and cross-sensor dependencies. The experimental results suggest that this component contributed mainly to the improvement in point prediction performance.

2. A posterior-motivated event-triggered graph update mechanism was introduced to refine dynamic sensor relationships more selectively. Instead of updating the graph freely at every time step, the proposed mechanism updated the dependency structure only when the trigger statistic suggested a potential change in sensor interactions. Under the considered settings, this design helped reduce unnecessary graph variation and supported a more structured description of dependency evolution during degradation.

Table 4. Ablation study of the proposed components in scenario I.

	RMSE				Score				MAPE			
Approach	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
GLSTM	17.47	15.37	15.82	16.41	485	1148	1202	1650	33.24	35.59	22.65	37.75
GLSTM + ST	12.98	11.21	11.07	12.68	265	654	258	798	13.93	15.20	13.53	21.61
GLSTM + BETD	15.52	12.54	15.39	15.12	470	845	1534	1375	16.28	21.11	25.66	21.68
GLSTM + BETD+ ST	11.84	11.22	11.59	13.09	245	648	229	781	13.56	14.26	14.98	20.38
GLSTM + UDF	16.84	14.78	15.75	15.25	441	1054	1126	1368	29.03	28.81	21.58	30.66
GLSTM + ST + UDF	13.00	11.12	10.93	**12.60**	258	611	236	768	13.46	15.25	14.55	18.82
GLSTM + BETD + UDF	13.05	12.46	14.33	14.23	282	845	1283	1227	16.33	18.25	16.78	22.70
GLSTM + BETD + ST + UDF	**11.79**	**11.08**	**9.84**	12.63	**239**	**601**	**210**	**734**	**13.05**	**14.17**	**13.79**	**18.26**

Table 5. Ablation study of the proposed components in scenario II.

	RMSE				Score				MAPE			
Approach	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004	FD001	FD002	FD003	FD004
GLSTM	18.35	16.63	16.13	17.01	557	1346	1229	1759	43.40	49.34	30.16	45.20
GLSTM + ST	13.07	12.08	11.87	12.81	271	751	255	816	14.80	17.32	15.29	24.04
GLSTM + BETD	14.92	13.51	15.40	14.79	505	973	1541	1358	16.97	24.46	27.70	26.12
GLSTM + BETD + ST	11.68	12.13	11.44	12.76	279	735	230	763	20.43	15.27	16.99	21.66
GLSTM + UDF	17.17	15.08	15.51	15.37	999	1110	1072	1360	35.23	35.00	24.65	35.14
GLSTM + ST + UDF	13.06	11.79	11.74	12.65	263	690	235	786	19.76	17.25	16.49	23.81
GLSTM + BETD + UDF	13.47	13.95	14.58	14.52	337	1023	1316	1285	17.69	23.01	20.10	22.07
GLSTM + BETD + ST+ UDF	**11.95**	**11.77**	**10.14**	**12.84**	**270**	**664**	**218**	**799**	**16.22**	**14.69**	**14.41**	**18.67**

3. An uncertainty-aware fusion strategy was introduced for aggregating sensor-level predictions. By estimating predictive uncertainty at the sensor level and using it to modulate sensor contributions, the proposed fusion strategy improved the handling of noisy or inconsistent sensor signals. The results indicate that this component was particularly helpful for uncertainty-aware prediction and interval-based evaluation.
4. The experimental study showed complementary roles of the three components. The spatial–temporal module contributed primarily to representation learning and prediction accuracy, the event-triggered graph mechanism helped stabilize dependency modeling, and the uncertainty-aware fusion module improved uncertainty-related performance under the tested conditions. These results suggest that the overall benefit of the framework comes from the combination of these complementary functions rather than from any single module alone.

Overall, the proposed framework provides a structured approach for RUL prediction that jointly considers temporal dynamics, sensor interactions, and predictive uncertainty. Although the current validation was limited to the considered benchmark settings, the results indicate that the framework is a useful step toward more robust and uncertainty-aware prognostics for complex equipment.

5.2. Future work

The proposed framework shows advantages in modeling temporal degradation patterns, capturing sensor dependencies, and improving robustness under sensor disturbance or missing-data conditions. However, some limitations remain. First, the current validation is mainly based on benchmark datasets and simulated sensor disturbances, which may not fully reflect complex industrial scenarios. Future work will further evaluate the framework on real industrial datasets with stronger operating-condition variability, more complex degradation modes, and richer sensor corruption or missing-data patterns. Second, the graph update scheme still relies on predefined design choices and hyperparameters. Future work will explore more adaptive triggering criteria that can be learned from data or automatically calibrated under different operating conditions.

Acknowledgments

The authors gratefully acknowledge the support from the National Natural Science Foundation of China (No. 52506259), Natural Science Foundation of Shanghai Municipality (No. 24ZR1454800), and the State Key Laboratory of Mechanical System and Vibration (No. MSV202411).

Data availability statements

Code and data availability are provided on request to the corresponding author.

Conflict of 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.

Author contributions

Zifei Xu 0000-0003-2661-517X
Conceptualization (lead), Data curation (equal), Formal analysis (equal), Funding acquisition (lead), Investigation (lead), Methodology (lead), Project administration (equal), Writing – original draft (lead)

Musa Bashir
Formal analysis (equal), Software (equal), Writing – review & editing (supporting)

Qiang Zhang
Methodology (supporting), Writing – original draft (supporting)

Michael Beer
Supervision (equal), Writing – review & editing (equal)

Jin Wang 0000-0003-4646-9106
Data curation (equal), Supervision (equal), Writing – review & editing (equal)

Zaili Yang
Supervision (equal), Writing – review & editing (equal)

Appendices

Appendix A. Training configuration

All experiments were implemented in PyTorch and trained on an NVIDIA RTX A5500 GPU. A fixed random seed of 42 was used to ensure reproducibility. For each dataset, the training set was split at the engine level, with 70% used for training and 30% for validation. The models were optimized using the AdamW optimizer with an initial learning rate of 3×10^{-4}, a batch size of 64, and a weight decay of 1×10^{-2}. Gradient clipping with a maximum norm of 1.0 was applied to stabilize training. The maximum number of training epochs was set to 1000.

An adaptive two-stage training strategy was employed. In the first stage, the model was warmed up using only the MSE objective, while the log-variance head was frozen. If the validation MSE did not improve by at least 1×10^{-4} for 8 consecutive epochs after a minimum of 5 warm-up epochs, or if the warm-up stage reached 500 epochs, the training objective was switched to the negative log-likelihood (NLL) loss. A 20-epoch ramp-up schedule was further used to smoothly transition from MSE-based optimization to NLL-based optimization. Early stopping was then applied with a patience of 50 epochs and a minimum improvement threshold of 5×10^{-5}.

For uncertainty-aware inference, MC dropout was adopted with 30 stochastic forward passes. The dropout rate was set to 0.1 during training and inference.

Appendix B. Pseudocode

Algorithm 1. Training procedure.

Require: $\mathcal{D}_{tr}$, $\mathcal{D}_{val}$, f_θ, $\mathcal{O}$, $E_{\max}$, P
Require: Warm-up settings $\mathcal{A}$
Ensure: Best parameters $\theta^\star$

Initialize $\theta, \theta^\star, m^\star \leftarrow +\infty$
useNLL $\leftarrow$ False; $c_{\text{pl}} \leftarrow 0$; $c_{\text{es}} \leftarrow 0$
if $\mathcal{A}$. freeze_logvar_head **then**
    Freeze variance head
end if
for $e = 1$ to $E_{\max}$ **do**
    Set loss schedule:

$$\mathcal{L} = \begin{cases} \mathcal{L}_{\text{MSE}}, & \text{useNLL} = \text{False} \\ (1-\alpha_e)\,\mathcal{L}_{\text{MSE}} + \alpha_e \mathcal{L}_{\text{NLL}}, & \text{useNLL} = \text{True} \end{cases}$$

    for all $(x, y, m, \ell) \in \mathcal{D}_{tr}$ **do**
        Compute sensor-wise outputs $(\mu_s, \sigma_s^2) \leftarrow f_\theta\,(x, m)$
        Update θ using $\mathcal{L}$
    end for
    Compute validation metrics M_e, N_e, and V_e on $\mathcal{D}_{val}$
    if $\neg$ useNLL **then**
        Update warm-up plateau counter using M_e
        if warm-up stopping criterion satisfied **then**
         useNLL $\leftarrow$ True; $e_{\text{switch}} \leftarrow e$
        if $\mathcal{A}$. freeze_logvar_head **then**
            Unfreeze variance head
        end if
        Reset $m^\star$ and c_{es}
        end if
    else
        Select validation metric m_e
        if $m_e < m^* - \delta$ **then**
            $m^* \leftarrow m_e; \theta^\star \leftarrow \theta; c_{\text{es}} \leftarrow 0$
        else if $e \geqslant e_{\text{switch}} + \max(1, \mathcal{A}.\textit{ramp\_epochs})$ then
        $c_{\text{es}} \leftarrow c_{\text{es}} + 1$
        if $c_{\text{es}} \geqslant P$ **then**
            break
         end if
        end if
    end if
    end for
Load $\theta^\star$

Algorithm 2. Uncertainty-aware dynamic fusion (UDF).

Require: Sensor-wise predictions $\mu_{t,i}^{(k)}$ and variances $\sigma_{t,i}^{2(k)}$, for $i = 1, \ldots, D$, $k = 1, \ldots, K$, and $t = 1, \ldots, T$; mask m_t; coefficients β and ε

Ensure: Fused mean $\hat{y}_t$ and variance v_t for each valid time step t

for $t = 1$ to T **do**

  for $i = 1$ to D **do**

    $\bar{\mu}_{t,i} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \mu_{t,i}^{(k)}$

    $a_{t,i} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \sigma_{t,i}^{2(k)}$

    $e_{t,i} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \left(\mu_{t,i}^{(k)} - \bar{\mu}_{t,i} \right)^2$

    $u_{t,i} \leftarrow a_{t,i} + \beta e_{t,i}$

    $p_{t,i} \leftarrow (u_{t,i} + \varepsilon)^{-1}$

  end for

  $\omega_{t,i} \leftarrow \frac{p_{t,i}}{\sum_{j=1}^{D} p_{t,j}}, \ i = 1, \ldots, D$

  $\hat{y}_t \leftarrow \sum_{i=1}^{D} \omega_{t,i} \bar{\mu}_{t,i}$

  $v_t \leftarrow \sum_{i=1}^{D} \omega_{t,i} \left(u_{t,i} + \bar{\mu}_{t,i}^2 \right) - \hat{y}_t^2$

  if $m_t = 0$ **then**

    $(\hat{y}_t, v_t) \leftarrow$ invalid

  end if

end for

Appendix C. Model hyper-parameters setting

Table 6. Main configuration of the proposed model.

Category	Setting	Value
Input setting	Number of sensors	14
Input setting	Sequence form	Full-sequence
Representation learning	Embedding dimension	64
Temporal modeling	Temporal kernel size	3
Temporal modeling	Dilation rates	1, 2, 4, 8
Temporal modeling	Depth per dilation	1
Spatial modeling	Attention heads	4
Spatial modeling	Feed-forward expansion factor	2
Recurrent modeling	GLSTM hidden dimension	128
Graph learning	Dynamic adjacency update	Enabled
Graph learning	Low-rank update rank	8
Graph learning	Residual scaling factor	0.10
Graph adaptation	Gate-on-update mechanism	Enabled
Graph adaptation	Gate slope	10
Uncertainty quantification	MC dropout passes	50
Uncertainty quantification	Training dropout rate	0.1
Uncertainty quantification	Inference dropout rate	0.1
Optimization	Optimizer	AdamW
Optimization	Initial learning rate	3×10^{-4}
Optimization	Weight decay	1×10^{-2}
Optimization	Batch size	64
Optimization	Maximum epochs	1000
Optimization	Gradient clipping norm	1.0
Early stopping	Patience	50

Table 7. Hyperparameter settings of the BETD module.

Component	Hyperparameter	Value
Prior change belief	prior probability of change π	0.01
Evidence model	Mean shift under change	2.5
Evidence normalization	EMA update rate	0.05
Triggering rule	Activation threshold D_{on}	0.7
Triggering rule	Deactivation threshold D_{on}	0.3
Triggering rule	Sustain steps	3
Graph update	Update rate η	0.5

Table 8. Hyperparameter settings of the UDF.

Component	Hyperparameter	Value
Uncertainty composition	Epistemic balancing coefficient β	1.0
Numerical stability	Small constant ε	1×10^{-8}
MC uncertainty estimation	Number of MC samples K	100

Table 9. Parameter breakdown of the proposed model.

Component	Trainable parameters
Input embedding layer	1,152
Temporal feature extractor	17 792
Sensor attention module	64 064
Feature fusion block	17 024
GLSTM with BETD	115 745
Prediction heads	16 642
Total	232 419

Table 10. Sensitivity analysis of key hyperparameters across all FD conditions using RMSE metric.

$\eta \backslash D_{on}$	0.60: $\beta \sim [0.5 \sim 2.0]$	0.70 $\beta \sim [0.5 \sim 2.0]$	0.80 $\beta \sim [0.5 \sim 2.0]$	0.90 $\beta \sim [0.5 \sim 2.0]$
0.2	[11.668, 11.834]	[11.571, 11.660]	[11.627, 11.782]	[12.523, 12.688]
0.5	[11.298, 11.390]	[11.175, 11.378]	[11.250, 11.348]	[12.160, 12.290]
0.8	[11.295, 11.458]	[11.194, 11.312]	[11.227, 11.443]	[12.132, 12.333]

Appendix D. Inference speed

Table 11. Inference-time on the four CMAPSS subsets.

Metric	FD001	FD002	FD003	FD004
Num of engines	100	259	100	248
Total time (s)	37.95	109.34	56.91	119.68
Average speed (ms)	379.55	442.16	569.16	482.59

References

Amin A, Bibo A, Panyam M and Tallapragada P 2023 An uncertainty-aware health monitoring model for wind turbine drivetrains based on Bayesian neural network *IFAC-PapersOnLine* **56** 235–40

Cheng Y, Qv J, Feng K and Han T 2024 A Bayesian adversarial probsparse Transformer model for long-term remaining useful life prediction *Reliab. Eng. Syst. Saf.* **248** 110188

Costa E A, Rebello C M, Santana V V and Nogueira I B R, & coauthors 2024 Physics-informed neural network uncertainty assessment through Bayesian inference *IFAC-PapersOnline* vol 58 pp 652–7 (available at: https://skoge.folk.ntnu.no/prost/proceedings/adchem-2024/files/0101.pdf)

Du J, Song L, Gui X, Zhang J, Guo L and Li X 2024 Remaining useful life prediction under variable operating conditions via multisource adversarial domain adaptation networks *Appl. Soft. Comput.* **161** 111717

Feng K, Ji J C, Ni Q and Beer M 2023a A review of vibration-based gear wear monitoring and prediction techniques *Mech. Syst. Signal Process.* **182** 109605

Feng K, Ji J C, Zhang Y, Ni Q, Liu Z and Beer M 2023b Digital twin-driven intelligent assessment of gear surface degradation *Mech. Syst. Signal Proc.* **186** 109896

Guo J, Song Y, Wang Z, Ma T, Xiao Y and Xu Z 2026 A hybrid data-driven prognostic scheme based on unsupervised health indicator construction and random-effects Wiener process *Comput. Ind. Eng.* **212** 111706

Guo Y, Sun Y, Si Q, Guo X and Chen N 2025 Probabilistic risk assessment of civil aircraft associated failures under condition-based maintenance *Reliab. Eng. Syst. Saf.* **253** 110550

Han T and Li Y-F 2022 Out-of-distribution detection-assisted trustworthy machinery fault diagnosis approach with uncertainty-aware deep ensembles *Reliab. Eng. Syst. Saf.* **226** 108648

Hervé de Beaulieu M, Jha M S, Garnier H and Cerbah F 2024 Remaining useful life prediction based on physics-informed data augmentation *Reliab. Eng. Syst. Saf.* **252** 110451

Jiang F, Hou X and Xia M 2025 Spatio-temporal attention-based hidden physics-informed neural network for remaining useful life prediction *Adv. Eng. Inf.* **63** 102958

Karatzinis G D, Boutalis Y S and Van Vaerenbergh S 2024 Aircraft engine remaining useful life prediction: a comparison study of kernel adaptive filtering architectures *Mech. Syst. Signal Proc.* **218** 111551

Keshun Y, Guangqi Q and Yingkui G 2024 Optimizing prior distribution parameters for probabilistic prediction of remaining useful life using deep learning *Reliab. Eng. Syst. Saf.* **242** 109793

Laflamme S *et al* 2026 Roadmap: Integrating Artificial Intelligence in Structural Health Monitoring Systems (*Meas. Sci. Technol.*) **37** 103001

Le Xuan Q, Munderloh M and Ostermann J 2024 Self-supervised domain adaptation for machinery remaining useful life prediction *Reliab. Eng. Syst. Saf.* **250** 110296

Lei Y 2016 *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery* (Butterworth-Heinemann) (https://doi.org/10.1016/c2016-0-00367-4)

Lei Y, Li N and Li X 2022 *Big Data-Driven Intelligent Fault Diagnosis and Prognosis for Mechanical Systems* (Springer) (https://doi.org/10.1007/978-981-16-9131-7)

Li D, Chen J, Huang R, Chen Z and Li W 2024 Sensor-aware CapsNet: towards trustworthy multisensory fusion for remaining useful life prediction *J. Manuf. Syst.* **72** 26–37

Li H, Deng Z-M, Golilarz N A and Guedes Soares C 2021 Reliability analysis of the main drive system of a CNC machine tool including early failures *Reliab. Eng. Syst. Saf.* **215** 107846

Li H, Ding Y and Guedes Soares C 2026a An intelligent failure mode identification model for wind turbines *J. Offshore Mech. Arct. Eng.* **148** 021701

Li H, Ding Y, Sun Y, Xie M and Guedes Soares C 2025 An intelligent failure feature learning method for failure and maintenance data management of wind turbines *Reliab. Eng. Syst. Saf.* **261** 111113

Li H, Teixeira A P and Soares C G 2020 A two-stage failure mode and effect analysis of offshore wind turbines *Renew. Ener.* **162** 1438–61

Li M, Xu Z, Li S, Kikuchi Y, Dong Y, Gryllias K C, Baraldi P, Zio E and Carroll J 2026b Health prognostics and maintenance decision-making for wind energy: a comprehensive overview *Renew. Sust. Ener. Rev.* **226** 116269

Liao X, Chen S, Wen P and Zhao S 2023 Remaining useful life with self-attention assisted physics-informed neural network *Adv. Eng. Inf.* **58** 102195

Liu K, Shang Y, Ouyang Q and Widanage W D 2021 A data-driven approach with uncertainty quantification for predicting future capacities and remaining useful life of lithium-ion battery *IEEE Trans. Ind. Electron.* **68** 3170–80

Liu L, Song X and Zhou Z 2022 Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture *Reliab. Eng. Syst. Saf.* **221** 108330

Liu Y, Hou B, Ahmed M, Mao Z, Feng J and Chen Z 2024 A hybrid deep learning approach for remaining useful life prediction of lithium-ion batteries based on discharging fragments *Appl. Ener.* **358** 122555

Lu X, Jiang Q, Shen Y, Lin X, Xu F and Zhu Q 2024 Enhanced residual convolutional domain adaptation network with CBAM for RUL prediction of cross-machine rolling bearing *Reliab. Eng. Syst. Saf.* **245** 109976

Luo Y, Ning Q, Chen B, Zhou X and Huang L 2024 STAGNN: a spatial-temporal attention graph neural network for network traffic prediction *Int. J. Commun. Netw. Distrib. Syst.* **30** 413–32

Meng H, Geng M and Han T 2023 Long short-term memory network with Bayesian optimization for health prognostics of lithium-ion batteries based on partial incremental capacity analysis *Reliab. Eng. Syst. Saf.* **236** 109288

Mukhoti J, Gal Y and Kendall A 2024 Aleatoric–epistemic uncertainty balancing in multi-task neural networks (2409.XXXXX)

Nejjar I, Geissmann F, Zhao M, Taal C and Fink O 2024 Domain adaptation via alignment of operation profile for remaining useful lifetime prediction *Reliab. Eng. Syst. Saf.* **242** 109718

Nemani V P, Lu H, Thelen A, Hu C and Zimmerman A T 2022 Ensembles of probabilistic LSTM predictors and correctors for bearing prognostics using industrial standards *Neurocomputing* **491** 575–96

Nguyen K T P, Medjaher K and Gogu C 2022 Probabilistic deep learning methodology for uncertainty quantification of remaining useful lifetime of multi-component systems *Reliab. Eng. Syst. Saf.* **222** 108383

Ni Q, Ji J and Feng K 2022 Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network *IEEE Trans. Ind. Inform.* **3203** 1301–11

Ovadia Y, Fertig E, Ren J, Nado Z, Sculley D, Nowozin S, Dillon J V and Lakshminarayanan B 2019 Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift *Advances in Neural Information Processing Systems* **32** 13991–4002

Pan J, Sun B, Wu Z, Yi Z, Feng Q, Ren Y and Wang Z 2024 Probabilistic remaining useful life prediction without lifetime labels: a Bayesian deep learning and stochastic process fusion method *Reliab. Eng. Syst. Saf.* **250** 110313

Peng W, Ye Z S and Chen N 2020 Bayesian deep-learning-based health prognostics toward prognostics uncertainty *IEEE Trans. Ind. Electron.* **67** 2283–93

Qian Q, Zhou J, Hou B, Wang J, Sheng H and Zhang J 2026 Dimension-mismatched adversarial network: a new feature distribution adaptation method for rolling bearing RUL prediction *Reliab. Eng. Syst. Saf.* **271** 112269

Qiu G, Ye B, Gu Y, Huang P, Li H and Xu Z 2025 A novel bearing remaining useful life prediction methodology with slope-based change point detection and WOA-attention-BiLSTM model *IEEE Sens. J.* **25** 10417–31

Sedehi O, Papadimitriou C and Katafygiotis L S 2022 Hierarchical Bayesian uncertainty quantification of finite element models using modal statistical information *Mech. Syst. Signal Proc.* **179** 109296

Shi Y, Wei P, Feng K, Feng D-C and Beer M 2025 A survey on machine learning approaches for uncertainty quantification of engineering systems *Mach. Learn. Comput. Sci. Eng.* **1** 11

Soualhi M, Nguyen K T P, Medjaher K, Nejjari F, Puig V, Blesa J, Quevedo J and Marlasca F 2023 Dealing with prognostics uncertainties: combination of direct and recursive remaining useful life estimations *Comput. Ind.* **144** 103766

Tian H, Mi J, Tong S and Li Y F 2026 Uncertainty-weighted with gradient-based to re-weight domain generalization for remaining useful life prediction of rotating machinery under unseen conditions *Reliab. Eng. Syst. Saf.* **267** 111810

Xu Z, Bashir M, Liu Q, Miao Z, Wang X, Wang J and Ekere N 2023 A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model *Comput. Ind. Eng.* **176** 108999

Xu Z, Zhao K, Zhang W, Miao W, Sun K, Wang J and Bashir M 2025 Collaborative and trustworthy fault diagnosis for mechanical systems based on probabilistic neural network with decision-level information fusion *J. Ind. Inf. Integr.* **46** 100854

Yang L, Wang J and Xie M 2025 Graph-based reliability evaluation of a reconfigurable multi-stage system using sequential unconnected path sets *Reliab. Eng. Syst. Saf.* **261** 111093

Yin C, Sun T, Wu H and Dong Y 2026 Trustworthy multistep-ahead remaining useful life prediction for rolling bearings with limited data *Reliab. Eng. Syst. Saf.* **267** 111902

Zhan Y, Kong Z, Wang Z, Jin X and Xu Z 2024 Remaining useful life prediction with uncertainty quantification based on multi-distribution fusion structure *Reliab. Eng. Syst. Saf.* **251** 110383

Zhang C, Gong D and Xue G 2025 An uncertainty-incorporated active data diffusion learning framework for few-shot equipment RUL prediction *Reliab. Eng. Syst. Saf.* **254** 110632

Zhang J X, Zhang J L, Zhang Z X, Li T M and Si X S 2024 Remaining useful life prediction for stochastic degrading devices incorporating quantization *Reliab. Eng. Syst. Saf.* **250** 110223

Zhang Y, Li Y, Wei X and Jia L 2020 Adaptive spatio-temporal graph convolutional neural network for remaining useful life estimation *Proc. Int. Joint Conf. on Neural Networks* (https://doi.org/10.1109/IJCNN48605.2020.9206739)

Zhao J, Wang Z, Ren W, Li Y, Chen Z, Dong L, Fan X and Zhang X 2026 Remaining useful life prediction of rolling bearings considering sensor degradation *Mech. Syst. Signal Proc.* **244** 113753

Zhou J, Luo J, Qi J and Qin Y 2026 Knowledge library network for non-exemplar incremental remaining useful life prediction *IEEE ASME Trans. Mechatr.* 1–11

Zhou J, Yang J, Qian Q and Qin Y 2024 A comprehensive survey of machine remaining useful life prediction approaches based on pattern recognition: taxonomy and challenges *Measurement Science and Technology* vol 35 (Institute of Physics)

Zhou T, Zhang L, Han T, Droguett E L, Mosleh A and Chan F T S 2023 An uncertainty-informed framework for trustworthy fault diagnosis in safety-critical applications *Reliab. Eng. Syst. Saf.* **229** 108865

Zhu R, Peng W, Wang D and Huang C G 2022 Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data *Mech. Syst. Signal Proc.* **183** 109628