



ARTICLE

A Bayesian Tree-Structured Kernel Learning Approach for Gaussian Process Prognostics of Ship Propulsion Systems

Xinyu Jia^{1,2}, Qingjie Wang¹, Rongchuan Zhang³ and Heng Ouyang^{1,*}

¹State Key Laboratory of Smart Power Distribution Equipment and System, School of Mechanical Engineering, Hebei University of Technology, Tianjin, China

²Engineering Risk Analysis Group, Technische Universität München, Arcisstr. 21, München, Germany

³Wuhan Second Ship Design and Research Institute, Wuhan, China

*Corresponding Author: Heng Ouyang. Email: ouyangheng@hebut.edu.cn

Received: 20 April 2026; Accepted: 08 July 2026; Published: 23 October 2026

ABSTRACT: Accurate state assessment and remaining useful life (RUL) prediction for ship propulsion systems are critical in ensuring maritime safety, reducing maintenance costs, and enhancing transportation network resilience. Although Gaussian process methods are widely used for RUL prediction, existing models often rely on deterministic assumptions or fixed kernel functions, which struggle to fully capture the nonlinear, multi-scale, and stochastic characteristics of degradation in ship propulsion components. This paper proposes an automatic kernel selection framework based on Bayesian tree structures, integrated with Gaussian Process Regression (GPR), for probabilistic degradation modeling and RUL prediction of critical components. The method uses additive and multiplicative operators to combine interpretable base kernels, thereby constructing a composite kernel space. A Bayesian inference strategy is employed to search for kernel structures and estimate hyperparameters simultaneously. To efficiently explore the kernel space, a two-stage beam search algorithm using model evidence is designed, along with a validation-assisted optimization mechanism to improve prediction accuracy and uncertainty calibration. Experiments were conducted on the PRONOSTIA bearing dataset and a high-fidelity ship propulsion system simulation dataset. The results show that the proposed method achieves competitive prediction accuracy and provides more reliable uncertainty quantification than fixed-kernel GPR models, supporting its application in health monitoring and risk-aware predictive maintenance of ship propulsion systems.

KEYWORDS: Gaussian process regression; automatic kernel learning; Bayesian inference; remaining useful life; ship propulsion systems

1 Introduction

Maritime transportation systems constitute a vital component of modern transport infrastructure, whose resilience depends not only on fixed assets such as ports and waterways but also on the operational reliability of ships, which serve as mobile nodes within this network [1]. In practice, failures of key propulsion system components can lead to loss of propulsion, navigation disruption, channel blockage, and port operation delays, thereby cascading local faults into systemic impacts on the maritime logistics network [2]. Among these components, critical rotating elements—such as rolling bearings in shaft systems—operate for extended periods under variable rotational speeds, varying loads, and complex marine environments, where the degradation process exhibits pronounced nonlinearity, multi-scale coupling, and stochastic uncertainties [3]. Therefore, achieving accurate state estimation and probabilistic remaining useful life (RUL)

prediction of these critical components is of great significance for early risk identification, maintenance decision optimization, and ultimately enhancing the resilience of maritime transportation systems [4].

In recent years, data-driven methods for fault diagnosis and prognostics of marine systems have attracted increasing attention. Machine learning and deep learning have shown strong performance in RUL prediction, where recurrent neural networks, particularly long short-term memory (LSTM) models, are effective in capturing temporal dependencies, and convolutional neural networks (CNNs) are commonly used for automatic feature extraction from raw sensor data [5,6]. However, RUL prediction with reliable uncertainty quantification remains limited in marine machinery, while the transition from corrective to proactive maintenance has become increasingly important [7]. Multimodal data fusion integrating vibration, temperature, and acoustic emission has shown advantages in fault diagnosis of rotating machinery [8]. For bearings in propulsion shafting, existing studies employ signal decomposition with neural networks or ensemble empirical mode decomposition with Kalman filtering for degradation tracking [9]. In addition, deep learning models such as multi-layer perceptron (MLP) and bidirectional long short-term memory model (BiLSTM) have been applied to construct health indicators and capture temporal dependencies, demonstrating strong performance in bearing RUL prediction [10]. The Mar-RUL approach combines Markov models, one-dimensional convolutional neural network (1D-CNN), and LSTM for turbocharger RUL prediction [11]. A domain adversarial neural network (DANN)-based two-stage framework mitigates cross-condition prediction bias [12]. In addition, domain adaptation methods and digital twin technologies have been increasingly adopted to improve cross-condition generalization and real-time monitoring capability in complex engineering systems, where digital twin frameworks further enable real-time state estimation and system-level fault tracking through data-model integration [13]. Digital twins combined with adaptive extended Kalman filtering enable real-time severity estimation of propeller and shafting faults [14,15]. For vessel electric motors, an integrated machine learning and probabilistic degradation modeling approach has been proposed to predict RUL under real operational profiles [16,17]. Beyond prognostics, data-driven surrogate modeling has demonstrated broader value in predicting critical mechanical properties, for instance, in forecasting slider track peeling strength as a function of process parameters [18].

Notwithstanding these advances, most existing methods are deterministic and lack reliable uncertainty estimation, which is crucial for risk-aware maintenance decisions. Therefore, probabilistic modeling approaches are needed for robust prognostics in marine systems [19,20]. The importance of uncertainty extends into related areas such as damage identification, where accounting for uncertainties in model errors, measurement noise, and even algorithmic parameters like regularization is critical for reliable solutions [21]. Uncertainty-aware sensor placement has also been considered in model updating [22]. Bayesian approaches provide a principled framework for uncertainty quantification and have been widely applied in structural system identification and model updating [23–25]. In marine engineering applications, Gaussian process regression (GPR) combined with Bayesian inference has been successfully applied to failure modeling of ship engines, demonstrating the effectiveness of probabilistic approaches for health state assessment [26,27]. However, many Bayesian prognostic methods rely on sampling-based techniques, which are computationally expensive and difficult to scale.

GPR naturally provides predictive mean and variance, making it suitable for probabilistic prognostics [28,29]. It offers a principled Bayesian nonparametric framework for uncertainty-aware modeling and has been widely applied in reliability engineering and prognostics applications. However, GPR performance depends heavily on kernel selection. Conventional kernel functions, such as squared exponential and Matern kernels, are often insufficient to capture non-stationary and multi-scale degradation behaviors in marine propulsion systems. This undermines the reliability of resilience-oriented decisions. To address this issue,

automatic kernel discovery methods have been proposed. Compositional kernel learning approaches construct complex kernels by combining simple base kernels through algebraic operations. Duvenaud et al. [30] pioneered compositional kernel search using addition and multiplication of base kernels. Subsequently, Kuok et al. [31] proposed a Bayesian optimal kernel selection approach for seismic attenuation modeling, wherein the optimal kernel is selected from a candidate pool by evaluating its plausibility using Bayesian inference. Li et al. [32] developed a self-adaptive GPR approach that integrates tree-structured composite kernel optimization with gradient-based hyperparameter tuning to predict temperature fluctuations in T-junction pipes of sodium-cooled fast reactors. Sun et al. [33] introduced differentiable neural kernel networks. Nonetheless, existing approaches still face challenges in balancing prediction accuracy, uncertainty reliability, and computational efficiency.

Despite these advances, important gaps remain. Most current GP-based prognostic models rely on fixed kernel structures or heuristic kernel construction strategies, which limit their ability to capture non-stationary and multi-scale degradation behaviors [30,34]. Although compositional kernel learning improves model flexibility, existing methods rarely combine structured kernel search, Bayesian evidence-based model selection, and validation-based refinement in a unified and efficient framework. Therefore, a principled kernel learning approach is still needed to improve predictive accuracy, uncertainty reliability, and model interpretability for complex degradation systems.

To address these gaps, this paper develops a Bayesian, tree-structured automatic kernel learning framework for probabilistic degradation modeling of ship propulsion components. The proposed framework constructs compositional kernels through a constrained grammar, allowing candidate kernel structures to remain interpretable while capturing complex degradation patterns. Kernel structures are then evaluated using Bayesian evidence, so that model selection is not solely driven by prediction error but also accounts for model complexity. A validation-assisted refinement step is further introduced to improve the stability of prediction and uncertainty estimation. The framework is evaluated on the PRONOSTIA bearing dataset and a high-fidelity ship propulsion simulation dataset. The results show that the proposed method achieves better prediction accuracy and more reliable uncertainty estimation than fixed-kernel GPR models, supporting its application in health monitoring and predictive maintenance of ship propulsion systems.

The remainder of this paper is organized as follows. [Section 2](#) presents the proposed framework. [Section 3](#) applies the framework to bearing RUL prediction. [Section 4](#) extends the framework to ship propulsion system state prediction. [Section 5](#) concludes the paper.

2 Proposed Methodology

This section presents the proposed Bayesian tree-structured GPR framework for probabilistic degradation modeling and remaining useful life prediction of ship propulsion systems. The framework integrates data preprocessing, automatic kernel structure learning, Gaussian process modeling, probabilistic prediction, and performance evaluation in a unified workflow. The overall procedure of the proposed method is illustrated in [Fig. 1](#), which provides a system-level view of the complete modeling pipeline from raw data to final health assessment and RUL prediction.

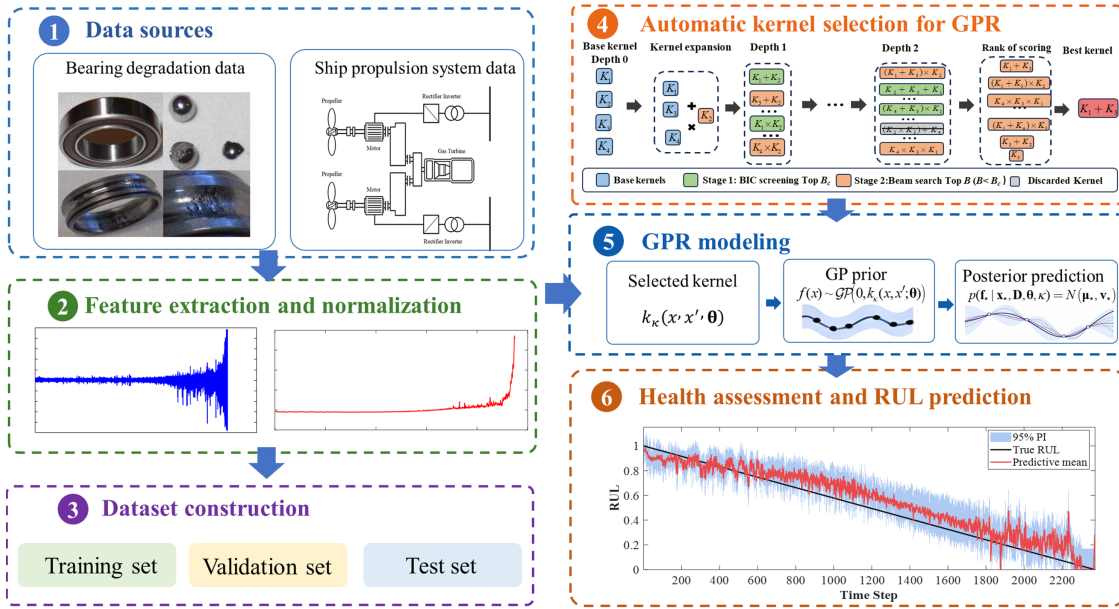


Figure 1: Overall workflow of the proposed Bayesian tree-structured GPR framework for probabilistic degradation modeling and remaining useful life prediction.

2.1 Problem Statement

This paper investigates the problems of probabilistic state estimation and degradation modeling for critical components of ship propulsion systems. These components operate under complex and time-varying conditions, and their degradation processes exhibit significant nonlinearity and randomness. Using multi-source sensor data (such as vibration, temperature, and operational parameters), this degradation modeling problem is formulated as a supervised regression task. Let the training dataset be defined as

$$\mathbf{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^d, y_i \in \mathbb{R} \quad (1)$$

where $\mathbf{x}_i$ denotes the input features extracted from sensor data and y_i represents the corresponding target variable. The objective is to learn a probabilistic mapping that provides accurate predictions and reliable uncertainty estimates for unseen inputs. The predictive distribution is modeled as:

$$y_* | \mathbf{x}_*, \mathbf{D} \sim \mathcal{N}(\mu(\mathbf{x}_*), \sigma^2(\mathbf{x}_*)) \quad (2)$$

where $\mu(\mathbf{x}_*)$ and $\sigma^2(\mathbf{x}_*)$ denote the predictive mean and predictive variance, respectively.

GPR is a Bayesian nonparametric framework that enables probabilistic prediction with uncertainty quantification [35,36]. The kernel function plays a central role in defining the covariance structure and modeling capability of GPR. In conventional approaches, kernels are typically specified manually, which may limit the flexibility of the model in representing complex data patterns. In this study, the kernel structure is inferred from data instead of being predefined.

2.2 Gaussian Process Regression

In GPR, the unknown latent function $f(x)$ is modeled as a stochastic process [29]. Specifically, a Gaussian process prior is assumed

$$f(x) \sim \mathcal{GP}(0, k_\kappa(x, x'; \boldsymbol{\theta})) \quad (3)$$

where $k_\kappa(\cdot, \cdot)$ is the covariance kernel, θ denotes continuous hyperparameters, and κ is indexed the kernel structure. The observed outputs are assumed to be corrupted by independent Gaussian noise

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_n^2) \quad (4)$$

where σ_n^2 represents the observation noise variance. Let $\mathbf{X} = [x_1, \dots, x_N]^T$, $\mathbf{y} = [y_1, \dots, y_N]^T$ denote the input matrix and output vector, respectively. The covariance matrix of the observations can then be written as

$$\mathbf{C}_\kappa(\theta) = \mathbf{K}_\kappa(\mathbf{X}, \mathbf{X}; \theta) + \sigma_n^2 \mathbf{I} \quad (5)$$

where $\mathbf{K}_\kappa(\mathbf{X}, \mathbf{X}; \theta)$ is the kernel matrix. Under the Gaussian process assumption, the marginal likelihood of the observed data becomes

$$p(\mathbf{y}|\theta, \kappa) = (2\pi)^{-N/2} |\mathbf{C}_\kappa|^{-1/2} \exp\left(-\frac{1}{2} \mathbf{y}^T \mathbf{C}_\kappa^{-1} \mathbf{y}\right) \quad (6)$$

For a new test input $\mathbf{x}_*$, define $\mathbf{k}_* = k_\kappa(\mathbf{X}, \mathbf{x}_*; \theta)$, $\mathbf{k}_{**} = k_\kappa(\mathbf{x}_*, \mathbf{x}_*; \theta)$. The predictive distribution of the latent function value $\mathbf{f}_* = f(\mathbf{x}_*)$ is given by

$$p(\mathbf{f}_*|\mathbf{x}_*, \mathbf{D}, \theta, \kappa) = N(\boldsymbol{\mu}_*, \mathbf{v}_*) \quad (7)$$

with predictive mean and variance

$$\boldsymbol{\mu}_* = \mathbf{k}_*^T \mathbf{C}_\kappa^{-1} \mathbf{y}, \quad \mathbf{v}_* = \mathbf{k}_{**} - \mathbf{k}_*^T \mathbf{C}_\kappa^{-1} \mathbf{k}_* \quad (8)$$

While this formulation provides both mean prediction and predictive uncertainty, its performance is governed by the choice of the kernel function. In the following, an automatic kernel discovery framework is introduced.

2.3 Kernel Structure Representation

Instead of specifying a predefined kernel family, this work constructs an explicitly defined compositional kernel space based on a predefined library of interpretable base kernels and algebraic composition rules. The proposed framework organizes the kernel search process in a Bayesian tree-structured manner, which facilitates systematic exploration of candidate kernels while maintaining model interpretability. Specifically, the kernel space is constructed from a predefined set of base kernels together with algebraic composition rules and can be viewed as the closure of the base kernel set under these operators. Representative approaches for compositional kernel discovery frameworks include the Automatic Statistician approach proposed by Duvenaud et al. [30] and tree-based genetic programming for GPR kernel construction [37]. In contrast, the proposed method focuses on efficient kernel exploration through structured search and a two-stage candidate screening strategy. It should be noted that the proposed approach does not search over an unrestricted kernel space. The search space is defined by a predefined base-kernel library and a compositional grammar, where candidate kernels are generated by recursively combining base kernels through addition and multiplication under structural constraints. More precisely, the grammar defines a constrained kernel space consisting of all kernels generated from the base kernel set via recursive application of the composition rules. Therefore, the method is more accurately described as a data-driven compositional kernel learning approach rather than an unconstrained kernel discovery method.

The construction begins with a library of interpretable base kernels

$$k_{base} = \{k_{RBF}, k_{Matern}, k_{RQ}, k_{lin}, k_{per}\} \quad (9)$$

including commonly used covariance functions such as the squared exponential (RBF) kernel, the Matern kernel, the rational quadratic kernel, the linear kernel, and the periodic kernel. These kernels capture fundamental functional patterns observed in engineering systems, such as smooth nonlinear trends, varying smoothness, multi-scale variability, long-term linear behavior, and periodic characteristics. The mathematical forms of these base kernels are summarized in Table 1 [29]. All base kernels can be used in either standard or ARD form, in the ARD version, each input dimension receives its own length scale ℓ_d , enabling automatic relevance determination.

Table 1: Mathematical expressions of base kernels.

Kernel	Mathematical Expression
Squared Exponential (RBF)	$k_{RBF}(x, x') = \sigma_f^2 \exp\left(-\frac{\ x - x'\ ^2}{2l^2}\right)$
Matern	$k_{Matern}(x, x') = \sigma_f^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\ x - x'\ }{l}\right)^\nu K_\nu\left(\frac{\sqrt{2\nu}\ x - x'\ }{l}\right)$
Rational Quadratic	$k_{RQ}(x, x') = \sigma_f^2 \left(1 + \frac{\ x - x'\ ^2}{2\alpha l^2}\right)^{-\alpha}$
Linear	$k_{Lin}(x, x') = \sigma_b^2 + \sigma_v^2 (x - c)(x' - c)$
Periodic	$k_{Periodic}(x, x') = \sigma_f^2 \exp\left(-\frac{2 \sin^2(\pi\ x - x'\ /p)}{l^2}\right)$

Note: For the Matern kernel, $\nu > 0$ controls the smoothness of the function; $\Gamma(\cdot)$ is the Gamma function; $K_\nu(\cdot)$ is the modified Bessel function of the second kind. In this paper, two specific cases are considered: $\nu = 3/2$ (Matern-3/2, denoted as Matern32) and $\nu = 5/2$ (Matern-5/2, denoted as Matern52).

More complex covariance functions can be constructed by combining base kernels through algebraic operations. Since kernel functions are closed under addition and multiplication, a compositional kernel grammar can be defined as follows [29]

$$\kappa ::= k_b \mid \kappa_1 + \kappa_2 \mid \kappa_1 \times \kappa_2, \quad k_b \in k_{base}. \quad (10)$$

This predefined grammar specifies the allowable ways in which base kernels can be composed, ensuring that the resulting kernel structures remain interpretable and within a constrained search space. A similar philosophy of composing complex representations from interpretable structural units has recently been adopted in structured panel encoding for multi-fidelity data fusion [38]. Under this formulation, each kernel is represented as a tree structure, leaf nodes represent base kernels, while internal nodes represent composition operators [37]. Additive combinations model the superposition of independent functional components, whereas multiplicative combinations capture interactions among different patterns.

However, the compositional kernel space grows rapidly with increasing tree depth. To control model complexity, this study considers the structural complexity of a compositional kernel from multiple aspects rather than using the number of tree nodes alone. The number of nodes reflects the overall size of the kernel expression, the number of hyperparameters characterizes the flexibility introduced by continuous kernel parameters, and the number of additive components reflects the degree of functional decomposition. In

addition, the maximum tree depth is explicitly controlled during candidate generation to avoid excessively nested kernel structures. These complexity-related quantities are incorporated into the Bayesian structural prior and the pruning rules used in the search process, ensuring that the exploration remains within a structured and tractable compositional kernel space.

2.4 Bayesian Kernel Structure Inference

The kernel learning problem is formulated within a Bayesian framework, where both kernel hyperparameters and kernel structures are treated probabilistically [39]. For a fixed kernel structure κ , all continuous parameters are collected in the vector

$$\Theta = [\theta^T, \sigma_n]^T \quad (11)$$

where θ denotes the set of kernel hyperparameters (e.g., length-scales, variance parameters, and periodicity parameters depending on the specific kernel composition) and σ_n represents the standard deviation of the independent Gaussian noise. Given the observed data, the posterior distribution of the parameters is given by [40,41]

$$p(\Theta|y, \kappa) = \frac{p(y|\Theta, \kappa) p(\Theta|\kappa)}{p(y|\kappa)} \quad (12)$$

In practice, the posterior is approximated by its maximum a posteriori (MAP) estimate

$$\hat{\Theta}_\kappa = \arg \max_{\Theta} (\log p(y|\Theta, \kappa) + \log p(\Theta|\kappa)) \quad (13)$$

The model evidence for a given kernel structure is defined as

$$p(y|\kappa) = \int p(y|\Theta, \kappa) p(\Theta|\kappa) d\Theta \quad (14)$$

which measures how well the structure explains the observed data. Because this integral is generally intractable, the Laplace approximation is applied around the MAP estimate. Let

$$\mathbf{H}_\kappa = -\nabla_{\Theta}^2 (\log p(y|\Theta, \kappa) + \log p(\Theta|\kappa))|_{\Theta=\hat{\Theta}_\kappa} \quad (15)$$

denote the negative Hessian of the log-posterior. The evidence can then be approximated as [42]

$$p(y|\kappa) \approx p(y|\hat{\Theta}_\kappa, \kappa) p(\hat{\Theta}_\kappa|\kappa) (2\pi)^{\frac{d_\kappa}{2}} |\mathbf{H}_\kappa|^{-\frac{1}{2}} \quad (16)$$

Taking the logarithm yields

$$\log p(y|\kappa) \approx \log p(y|\hat{\Theta}_\kappa, \kappa) + \log p(\hat{\Theta}_\kappa|\kappa) + \frac{d_\kappa}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{H}_\kappa| \quad (17)$$

The Laplace approximation provides a practical way to estimate the model evidence with moderate computational cost. Compared with sampling-based methods such as Markov chain Monte Carlo, the Laplace approximation is more computationally efficient and is therefore suitable for evaluating many candidate kernel structures. To encourage compact and interpretable kernel structures, the complexity of a compositional kernel is controlled by both a structural prior and explicit pruning constraints [43]. It should be noted that model complexity is not measured solely by the number of tree nodes. Instead, the structural prior jointly considers multiple complementary factors, including the number of nodes,

continuous hyperparameters, and additive operators. The node count captures the overall structural size of the kernel, while the number of hyperparameters reflects model flexibility and its potential to overfit. The additive-operator term penalizes excessive functional decomposition, which enlarges the function space through independent additive components. Multiplicative compositions are implicitly accounted for via the node and parameter counts, whereas additive operators are explicitly penalized due to their stronger effect on increasing model expressiveness. Accordingly, the complexity penalty $\Omega(\kappa)$ is defined as a weighted structural penalty:

$$\Omega(\kappa) = \lambda \mathbf{g}(\kappa) \quad (18)$$

where $\mathbf{g}(\kappa) = [N_{\text{nodes}}, N_{\text{params}}, N_{\text{add}}]^T$ denote the number of tree nodes, the number of continuous hyperparameters, and the number of additive operators, respectively. The coefficients $\lambda = [\lambda_1, \lambda_2, \lambda_3]$ control the strengths of the corresponding complexity penalties. Accordingly, the structural prior can be written as

$$\log p(\kappa) = -\Omega(\kappa). \quad (19)$$

Instead of incorporating tree depth as an additional prior term, we directly constrain the maximum depth by a predefined limit H_{max} during candidate generation. This design separates search-space control from model scoring, avoiding redundancy while effectively preventing excessively nested kernel structures. This formulation ensures that model complexity is not determined by node count alone, but by a combination of structural size, parameterization, and functional composition. Finally, the posterior over kernel structures is approximated as:

$$\log p(\kappa | \mathbf{D}) \propto \log p(y | \kappa) + \log p(\kappa) \quad (20)$$

The resulting objective defines a Bayesian score as the sum of the marginal likelihood and the structural prior.

$$\kappa^* = \arg \max_{\kappa} (\log p(y|\kappa) + \log p(\kappa)). \quad (21)$$

2.5 Evidence-Based Kernel Structure Search

To efficiently explore the large kernel structure space, a two-stage screening strategy is employed. Additional mechanisms are introduced to maintain computational tractability, as the number of possible kernel compositions grows rapidly with search depth. In the first stage, a fast approximation using the Bayesian Information Criterion (BIC) performs coarse filtering. For each candidate kernel κ , the BIC score is computed as

$$S_{\text{BIC}}(\kappa) = \log p(y | \hat{\Theta}_{\kappa}, \kappa) - \frac{d_{\kappa}}{2} \log N, \quad (22)$$

where $\hat{\Theta}_{\kappa}$ is the MAP estimate for the kernel, d_{κ} is the parameter dimension and N is the sample size.

In the second stage, the retained B_c candidates are evaluated using the Laplace-approximated model evidence together with the structural prior. Compared with the BIC approximation, the Laplace-approximated evidence provides a more accurate assessment of model fitness by incorporating both data likelihood and parameter uncertainty. The final Bayesian structure score for each kernel structure is defined as

$$S(\kappa) = \log p(y | \mathbf{X}, \kappa) + \log p(\kappa) \quad (23)$$

where $\log p(y|\mathbf{X}, \kappa)$ denotes the Laplace-approximated model evidence, and $\log p(\kappa)$ is the structural prior used to penalize overly complex kernel structures. Kernel structures are then expanded iteratively via a beam-search strategy. At each search depth, candidate kernels are generated from the current beam using additive and multiplicative composition operators. After BIC-based coarse screening and Bayesian fine evaluation, only the top B candidates, ranked by the Bayesian structure score, are retained as the beam for the next search depth. Here, B denotes the beam width, defined as the maximum number of candidate kernel structures retained at each search depth during beam search. It controls the trade-off between search diversity and computational cost: a small B may miss promising kernel structures, while a large B improves search diversity at the expense of increased computational cost. Therefore, the beam width provides a trade-off between search coverage and computational efficiency.

To prevent excessive growth of the search tree and reduce the search space, several structural pruning rules are applied during candidate expansion.

- Duplicate removal: Kernel structures generated by commutative operations are treated as equivalent and removed, for example, $k_1 + k_2$, $k_2 + k_1$, or $k_1 k_2$ and $k_2 k_1$.
- Depth constraint: Candidate kernels whose tree depth exceeds a predefined maximum are discarded to prevent excessively complex structures.
- Structural constraint: Kernels violating grammar rules (e.g., repeated nesting of identical operators) are eliminated.
- Early stopping based on evidence: Kernels that fail to improve the evidence score relative to their parent are pruned.

By integrating a compositional kernel representation, Bayesian evidence evaluation, and structured kernel search, the proposed framework provides an automatic and interpretable approach to kernel learning for GPR in engineering applications. Although the procedure efficiently identifies high-quality candidate kernels, selecting the final model based solely on Bayesian evidence may not always optimize predictive performance or uncertainty calibration. A validation-assisted refinement strategy is introduced in the next section. The overall beam-search-based kernel structure discovery process is illustrated in Fig. 2.

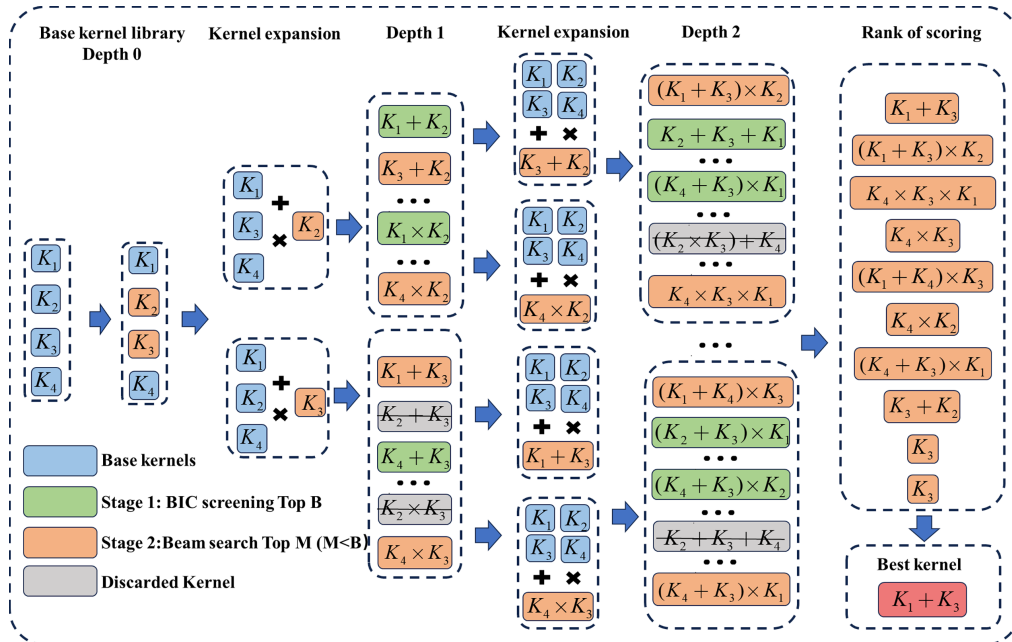


Figure 2: Two-stage beam-search framework for compositional kernel structure discovery.

2.6 Validation-Assisted Bayesian Model Selection

Although Bayesian evidence provides a principled criterion for model comparison, it may not distinguish sufficiently between structurally different kernels with similar marginal likelihoods. In general model selection practice, it is often beneficial to combine information criteria with out-of-sample validation metrics to obtain a more robust final model. Instead of selecting the kernel solely based on the highest Bayesian score, we first define a near-optimal candidate set and then apply a validation-assisted refinement step to improve predictive robustness. The proposed model selection procedure follows a Bayesian-first and validation-assisted strategy. During the kernel search stage, candidate structures are generated and ranked solely according to the Bayesian score, which consists of the Laplace-approximated model evidence and the structural prior. The validation set is not used to guide candidate generation, beam expansion, or hyperparameter optimization, ensuring that the search process remains fully consistent with Bayesian principles. After the search is completed, a near-optimal candidate set is constructed based on the Bayesian score. Specifically, let $S(\kappa^*)$ denote the highest Bayesian score. A candidate kernel κ is included in the near-optimal set if

$$S(\kappa) \geq S(\kappa^*) - \Delta \quad (24)$$

where $S(\kappa)$ is the Bayesian score, κ^* denotes the highest-scoring kernel structure, and $\Delta \geq 0$ is a predefined Bayesian-score tolerance. This parameter determines the size of the near-optimal candidate set, and kernels satisfying this condition are regarded as evidence-competitive models. The validation-assisted auxiliary score is then used only within this near-optimal set for final model selection.

Within this candidate set, predictive accuracy is evaluated using the root mean squared error (RMSE), while probabilistic consistency is assessed with the negative log-likelihood (NLL) and prediction interval coverage probability (PICP). Before computing the auxiliary score, RMSE and NLL are normalized within the near-optimal candidate set using min-max normalization to avoid scale dominance caused by different numerical ranges. The PICP term is computed as the absolute deviation from the nominal confidence level. These aspects are combined into an auxiliary score:

$$\text{AuxScore}(\kappa) = \alpha \cdot \text{RMSE}(\kappa) + \beta \cdot \text{PICP_err}(\kappa) + \gamma \cdot \text{NLL}(\kappa) \quad (25)$$

where PICP_err measures the deviation of PICP from the nominal confidence level, and α, β, γ are weighting coefficients. The weighting coefficients are determined based on preliminary empirical experiments to balance predictive accuracy and probabilistic reliability, and are kept fixed across all datasets and experiments to ensure fairness and avoid dataset-dependent tuning. The final kernel structure is selected as:

$$\kappa_{\text{final}} = \arg \min_{\kappa \in \kappa^*} \text{AuxScore}(\kappa) \quad (26)$$

This strategy ensures that the Bayesian score defines the primary selection criterion, while the validation-based auxiliary score acts as a tie-breaking mechanism among evidence-competitive models. As a result, the proposed approach improves predictive performance and uncertainty reliability while remaining consistent with Bayesian evidence maximization and without altering the underlying theoretical objective.

The overall procedure of the proposed Bayesian tree-structured kernel learning framework is summarized in Algorithm A1 ([Appendix A](#)).

3 Applications of Remaining Life Prediction for Marine Bearings

3.1 Experimental Datasets

During the actual service of marine propulsion shaft bearings, it is difficult to obtain complete and continuous full-life degradation data. This is mainly due to the complex operating environment, limited monitoring conditions, and high cost of long-term tracking. To verify the applicability of the proposed method to bearing RUL prediction, the PRONOSTIA accelerated bearing degradation dataset released in the IEEE PHM 2012 Challenge is adopted in this study [44]. This dataset has been widely used in rotating machinery prognostics and is highly representative in terms of degradation evolution and failure modes [12]. The PRONOSTIA platform consists of a rotating drive system, a loading system, and a measurement system, as shown in Fig. 3. During the experiment, the bearing operates under controlled rotational speed and radial load conditions, with the system collecting operating status through vibration and temperature sensors. Considering that bearings in ship propulsion shaft systems also exhibit time-evolving performance degradation characteristics during service, this dataset can be used to verify the effectiveness of the method in degradation modeling and life prediction tasks.

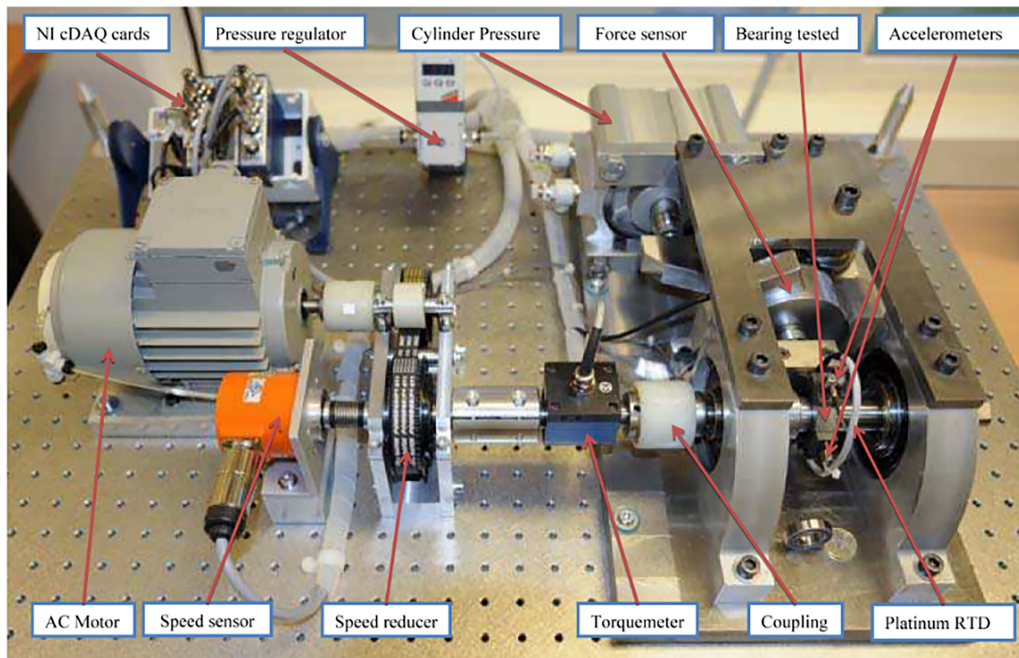


Figure 3: Structure of the PRONOSTIA experimental platform.

In this paper, data from Condition 1, corresponding to a rotational speed of 1800 rpm and a radial load of 4000 N, is selected for analysis. Although the original data provides both vibration and temperature information, only the horizontal vibration signal is used for modeling. To retain local temporal evolution information, a sliding window approach with a window length of 3 and a step size of 1 is used to construct the samples. In terms of features, time-frequency domain statistical features are extracted from each sliding window sample, resulting in 26-dimensional input variables. These include 14 time-domain features, such as mean, root mean square, peak value, and kurtosis, and 12 frequency-domain features, such as spectral energy, centroid frequency, and frequency standard deviation. This approach preserves degradation trend information while reducing the dimensionality of the original vibration data, thereby improving the stability of subsequent kernel function learning. To construct the remaining life label, a normalized remaining life

representation is adopted. Suppose a bearing's entire life consists of N sampling files, and the end position of the i sliding window is denoted as e_i . The label is defined as:

$$y_i = 1 - \frac{e_i - 1}{N - 1} = \frac{N - e_i}{N - 1} \quad (27)$$

The label gradually decreases from 1 to 0 as degradation progresses, representing the normalized RUL at the end of each window.

3.2 Automatic Kernel Selection Results for Bearing Degradation

Bearing degradation exhibits a global decreasing trend accompanied by local stochastic fluctuations. These characteristics introduce strong non-stationarity that cannot be captured by a single fixed kernel. To address these challenges, a compositional kernel search space is constructed based on a predefined set of base kernels, and candidate kernel structures are generated through a tree-structured expansion strategy. The kernel selection results in Table 2 show that, although some single kernels achieve high Bayesian scores, the best-performing models in terms of predictive accuracy and uncertainty calibration are primarily composite kernels, indicating that the bearing degradation process exhibits pronounced multi-component statistical characteristics.

Table 2: Top-ranked kernel structures and validation metrics for bearing degradation prediction.

Rank	Kernel	Bayes Score	RMSE (Val)	PICP (Val)	NLL (Val)
1	k_{MAT32}	1767.07	0.1133	0.156	25.69
2	k_{RQ}	1623.16	0.1055	0.157	40.25
3	$k_{\text{LIN}} + k_{\text{RQ}}$	1563.31	0.0808	0.690	0.812
4	k_{RBF}	1499.10	0.1131	0.143	0.994
5	$k_{\text{LIN}} + k_{\text{RQ}} + k_{\text{MAT52}}^*$	1391.65	0.0920	0.872	-0.738

Note: Negative NLL values result from the subtraction of constant baseline terms in the normalized log-likelihood; lower values still indicate better probabilistic fit. The asterisk (*) indicates the kernel structure finally selected by the proposed Bayesian-first, validation-assisted strategy for each prediction task.

As shown in Table 2, single kernels such as Matern32, RQ, and RBF achieve relatively high Bayesian scores, but their validation performance is unsatisfactory. These models exhibit low PICP values and large NLL values, indicating poor uncertainty calibration. In contrast, composite kernels demonstrate improved performance. The kernel $k_{\text{LIN}} + k_{\text{RQ}}$ achieves the lowest RMSE, while the full composite $k_{\text{LIN}} + k_{\text{RQ}} + k_{\text{MAT52}}$ provides the best probabilistic performance, with significantly higher PICP and lower NLL. The automatically identified kernel structure $k_{\text{LIN}} + k_{\text{RQ}} + k_{\text{MAT52}}$ reflects the multi-component nature of the degradation process. The linear kernel captures the global degradation trend, the Matern kernel models locally correlated variations, and the rational quadratic kernel accounts for multi-scale stochastic fluctuations. This additive composition enables the model to simultaneously represent deterministic trends and stochastic variations, thereby enhancing model flexibility while maintaining interpretability. Kernel choice mainly affects uncertainty quantification rather than point prediction accuracy. Single-kernel models tend to produce over-confident predictions due to variance collapse, whereas composite kernels provide well-calibrated predictive uncertainty. Therefore, the primary advantage of the proposed automatic kernel learning framework lies in constructing an appropriate covariance structure that reflects the intrinsic multi-scale and stochastic characteristics of bearing degradation, which is essential for reliable and risk-aware RUL prediction.

3.3 Predictive Performance and Uncertainty Analysis

Based on the determined optimal kernel structure, the normalized RUL prediction results for Bearing 1_3 are shown in Fig. 4. The predictive mean obtained from the composite kernel closely follows the true degradation trajectory across all stages. Uncertainty increases in the late degradation stage, and the prediction interval adapts by widening near end-of-life. In comparison, the single Matern32 kernel captures the general degradation trend but provides narrower prediction intervals, leading to reduced coverage performance and weaker uncertainty calibration.

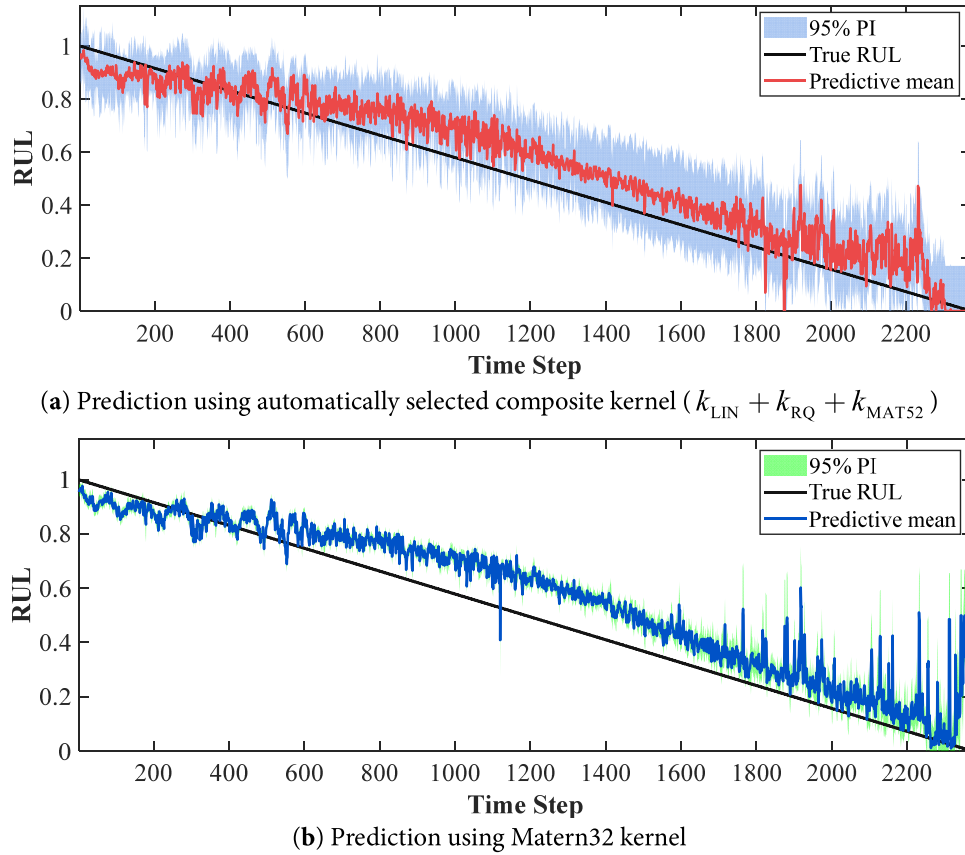


Figure 4: Bearing 1_3 RUL prediction results.

The improved performance of the composite kernel is mainly attributed to its ability to represent different degradation characteristics simultaneously. Specifically, the composite structure can capture the global degradation trend, local correlated variations, and multi-scale stochastic fluctuations, whereas the single Matern32 kernel provides a more limited covariance representation.

To further clarify the quantitative prediction performance, the numerical results corresponding to Fig. 5 are summarized in Table 3. RMSE is used to evaluate point prediction accuracy, PICP is used to measure the coverage probability of the prediction interval, and NLL is used to assess probabilistic calibration.

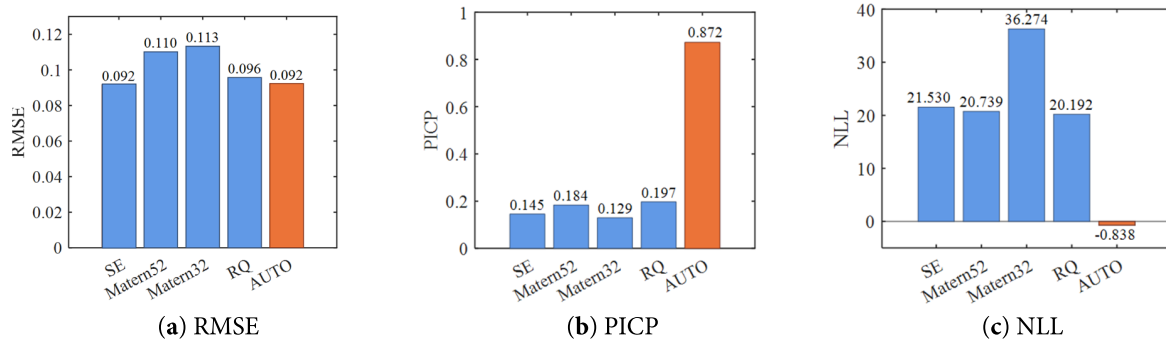


Figure 5: Comparison between fixed-kernel models and the automatically selected kernel. The blue bars denote fixed-kernel baselines, whereas the orange bar denotes the automatically selected kernel (AUTO). (a) RMSE for prediction accuracy, (b) PICP for uncertainty coverage, and (c) NLL for probabilistic calibration.

Table 3: Quantitative evaluation of prediction performance for Bearing 1_3 RUL prediction.

Kernel Configuration	RMSE	PICP	NLL	MAE
SE	0.092	0.145	21.530	0.07703
Matern52	0.110	0.184	20.739	0.09210
Matern32	0.113	0.129	36.724	0.09461
RQ	0.096	0.197	20.192	0.08038
AUTO	0.092	0.872	-0.838	0.05789

As shown in Table 3, the AUTO kernel achieves an RMSE of 0.092, which is comparable to the best fixed-kernel baseline. More importantly, its PICP reaches 0.872, which is substantially higher than those of the fixed-kernel models. In addition, the AUTO kernel obtains the lowest NLL value, indicating better probabilistic calibration. These results show that the proposed method not only maintains competitive point prediction accuracy but also provides more reliable uncertainty estimation.

Overall, the proposed Gaussian process model with automatic kernel learning achieves a better balance between prediction accuracy and uncertainty reliability. This advantage is particularly important for bearing RUL prediction in ship propulsion systems, where maintenance decisions require both accurate RUL estimates and reliable uncertainty assessment.

3.4 Comparison with Related Bearing RUL Prediction Studies

To compare the proposed method with related studies, representative bearing RUL prediction results on the PRONOSTIA/PHM2012 dataset are summarized in Table 4. Since different studies use different data partitions and experimental settings, the comparison is used as a contextual reference rather than a strict one-to-one benchmark.

As shown in Table 4, the proposed method achieves comparable prediction accuracy to existing deep learning and transfer learning methods. Although its RMSE is slightly higher, it provides explicit uncertainty quantification, which is not reported in the compared studies. This is important for risk-aware maintenance decisions, because practical RUL prediction requires both accurate point estimates and reliable uncertainty assessment.

Table 4: Comparison with related bearing RUL prediction studies.

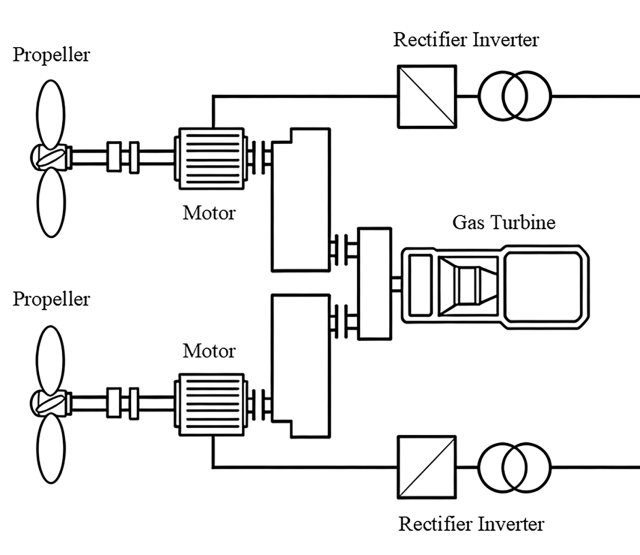
Study	Method	R ²	MAE	RMSE	Uncertainty Quantification
Ren et al. [12]	Enhanced DANN	0.939	0.0572	0.0732	No
Sun et al. [45]	CBAM-CNN-LSTM	0.943	0.0370	0.0690	No
Present study	Bayesian tree-structured GPR	0.944	0.0578	0.0921	Yes

Note: The results are taken from the corresponding reported Bearing 1_3 cases. Due to differences in experimental settings, the comparison is for reference only.

4 Application to Structural State Prediction of Marine Propulsion Systems

4.1 Experimental Data Description

To validate the applicability of the proposed Bayesian tree-structured automatic kernel learning method in complex engineering scenarios, a dataset generated from a high-fidelity numerical simulator of a naval propulsion system is adopted. The simulator models a frigate equipped with a CODLAG propulsion plant under various operating conditions and degradation states. The dataset is obtained from the UCI Condition-Based Maintenance of Naval Propulsion Plants repository [46,47]. At present, it can be downloaded from <https://cbm-anomaly-detection.smartlab.ws> [48]. The simulator has been calibrated using real ship data and is capable of reproducing both steady-state measurable variables and degradation processes of key components, including the gas turbine compressor, hull, and propeller. The structure of the CODLAG propulsion system is illustrated in Fig. 6.

**Figure 6:** Schematic of the CODLAG propulsion system.

The dataset consists of control inputs, sensor measurements, and degradation-related parameters. The system degradation is characterized by five degradation coefficients, whose physical meanings and ranges of values are summarized in Table 5. These degradation coefficients are generated by enumerating different combinations of degradation levels and control lever positions. A total of 455,109 samples is obtained, covering all possible steady-state operating conditions. For each parameter combination, the simulator is executed until a steady state is reached, and the corresponding observable variables are recorded. The outputs

are consistent with measurements available from onboard automation systems and include 25 input variables related to the control lever position, vessel speed, gas turbine operation, shaft dynamics, and propeller performance, as defined in the original UCI Condition-Based Maintenance of Naval Propulsion Plants dataset and related studies [46,47].

Table 5: Key CODLAG propulsion system component degradation coefficients and ranges.

Component	Symbol	Description	Range
Propeller	k_{Kt}	Thrust coefficient decay	0.9–1.0
Propeller	k_{Kq}	Torque coefficient increase	1.0–1.1
Hull	k_H	Hull resistance increases	1.0–1.2
Compressor	k_{Mc}	GTC degradation coefficient	0.95–1.0
Gas Turbine	k_{Mt}	GT degradation coefficient	0.975–1.0

During data preprocessing, invalid samples are removed, and feature reduction is performed to eliminate duplicated and highly correlated variables. Due to the large dataset size, stratified sampling is adopted based on degradation indicators and control lever positions. A subset of 2000 samples is selected. This subset preserves the distributional diversity of the full dataset and proportionally covers different degradation levels and operating conditions across all 25 input variables and 4 outputs. Using the full dataset directly for repeated GPR training and kernel search is computationally prohibitive, as GPR involves matrix inversion that is cubic in the number of training samples. The selected 2000 samples are then divided into training, validation, and test sets using a 70:15:15 ratio. The training set is used for kernel structure search and parameter estimation, the validation set for model selection, and the test set for evaluating generalization performance. Since the dataset consists of steady-state samples generated from independent operating conditions, no temporal dependency exists among samples. Therefore, the problem is formulated as a static mapping task rather than a time-series prediction or RUL estimation problem. Specifically, the objective is to learn a mapping from sensor measurements to degradation coefficients, enabling the estimation of degradation levels of key components, including the gas turbine, compressor, hull, and propeller. This facilitates structural health state assessment of the propulsion system.

4.2 Automatic Kernel Selection Results for Marine Propulsion State Prediction

To evaluate the adaptability of the proposed method in learning kernel structures for marine propulsion system state prediction, an automatic kernel search is conducted for four structural health indicators. Considering that the thrust coefficient decay and torque coefficient increase exhibit a linear relationship, this study focuses on modeling and predicting the thrust coefficient decay only. During the kernel search process, candidate kernel structures are progressively constructed from a predefined set of base kernels using a compositional grammar, where addition and multiplication operators are recursively applied. To improve search efficiency and ensure robust model selection, a two-stage screening strategy is adopted. Specifically, an initial coarse filtering is performed, followed by a refined evaluation based on Laplace-approximated marginal likelihood combined with structural priors. The final model selection follows a Bayesian-first, validation-assisted strategy, in which candidate kernels are primarily ranked according to their Bayesian scores, while validation metrics are used to assist discrimination among near-optimal candidates.

The candidate kernels are ranked according to their Bayesian scores, and the optimal kernel structures along with their corresponding performance metrics for each prediction task are summarized in Table 6. It

can be observed that the optimal kernel structures differ significantly across different target variables, indicating that distinct degradation indicators exhibit fundamentally different underlying covariance patterns. This further suggests that a fixed kernel function is insufficient to capture the diverse characteristics of all tasks.

Table 6: Top-3 kernels and evaluation metrics for each degradation coefficient.

Target	Rank	Kernel	Bayes Score	RMSE (Val)	PICP (Val)	NLL (Val)
k_{Mc}	1	$(k_{LIN} + k_{MAT32_{ARD}}) \times k_{RQ}^*$	1195.168	0.00121	0.92134	-5.45492
	2	$(k_{LIN} + k_{RQ}) \times k_{RBF_{ARD}}$	1135.887	0.00121	0.86516	-5.26210
	3	$(k_{LIN} \times k_{RQ}) + k_{MAT32_{ARD}}$	1074.156	0.00120	0.89513	-5.24095
k_H	1	$k_{LIN} \times k_{RQ}^*$	1167.028	0.00478	0.92509	-3.89155
	2	$(k_{LIN} \times k_{RQ}) + k_{MAT32_{ARD}}$	1162.734	0.00519	0.94007	-3.89420
	3	$(k_{LIN} \times k_{RBF_{ARD}}) + k_{RQ}$	1114.432	0.00503	0.92883	-3.86430
k_{Mt}	1	$(k_{LIN} + k_{RQ}) \times k_{RBF_{ARD}}^*$	1750.918	0.00020	0.98501	-6.83415
	2	$(k_{LIN} \times k_{RQ}) + k_{RBF_{ARD}}$	1648.324	0.00022	0.98501	-6.69000
	3	$(k_{LIN} + k_{RBF_{ARD}}) \times k_{RQ}$	1632.154	0.00022	0.92276	-6.70260
k_{Kq}	1	$k_{LIN} \times k_{RBF_{ARD}} \times k_{RQ}^*$	823.862	0.00171	0.97752	-5.09929
	2	$(k_{LIN} + k_{MAT32_{ARD}}) \times k_{RBF_{ARD}}$	824.383	0.00392	0.97003	-4.69400
	3	$(k_{LIN} \times k_{MAT32_{ARD}}) + k_{RQ}$	741.938	0.00354	0.97003	-4.65750

Note: The asterisk (*) indicates the final kernel selected by the Bayesian-first, validation-assisted strategy. The selected kernel is chosen from the near-optimal candidate set.

The optimal kernels can be grouped into two patterns. The first pattern corresponds to more complex composite kernels. This is observed in the k_{Mt} , k_{Mc} and k_{Kq} tasks. The optimal kernels are $k_{Mc} = (k_{LIN} + k_{MAT32_{ARD}}) \times k_{RQ}$, $k_{Mt} = (k_{LIN} + k_{RQ}) \times k_{RBF_{ARD}}$ and $k_{Kq} = k_{LIN} \times k_{RBF_{ARD}} \times k_{RQ}$. These kernels include linear components, multi-scale kernels, and ARD-based smooth kernels. They achieve strong performance on validation metrics. For example, the RMSE is 0.00171 for k_{Kq} and 0.00020 for k_{Mt} . This indicates that these degradation processes involve global trends, multi-scale variations, and input-dependent heterogeneity. The use of ARD suggests that input variables contribute unevenly and require anisotropic modeling. Notably, the optimal kernel for k_{Kq} adopts a fully multiplicative structure. This structure suggests that relying solely on the Bayesian score is insufficient, especially when candidate kernels have similar validation results. When Bayesian scores are close, the validation set metrics provide secondary validation, helping to make more reliable choices. It combines a linear-smooth interaction term with a multi-scale component. This is consistent with the multiplicative structures observed in other tasks, suggesting that the statistical behavior of this indicator is better described by coupled trend and fluctuation effects.

The second pattern corresponds to simpler kernels. For the k_H task, the optimal kernel $k_{LIN} \times k_{RQ}$ does not include ARD components but still achieves satisfactory performance. This indicates that hull degradation is mainly governed by a global trend with moderate scale variation and limited sensitivity to input dimensional differences. It is worth noting that some candidate kernels achieve similar validation errors but differ in Bayesian scores. This is observed in the k_{Kq} and k_{Mc} tasks. This suggests that validation metrics alone are insufficient. The Bayesian score provides additional discrimination by penalizing overly complex models. When Bayesian scores are close, validation metrics help distinguish near-optimal candidates. Therefore, the proposed strategy ensures both robustness and predictive reliability.

In summary, the proposed method can adaptively identify task-specific kernel structures. Complex processes require expressive kernels, while simpler processes favor compact ones. This reflects a consistent relationship between kernel complexity and system characteristics.

4.3 Prediction Results

To evaluate the effectiveness of the proposed Bayesian tree-structured automatic kernel learning framework, predictive performance was assessed on four degradation indicators. The results are presented in Figs. 7 and 8, where Fig. 7 illustrates prediction behavior and Fig. 8 provides quantitative evaluation metrics.

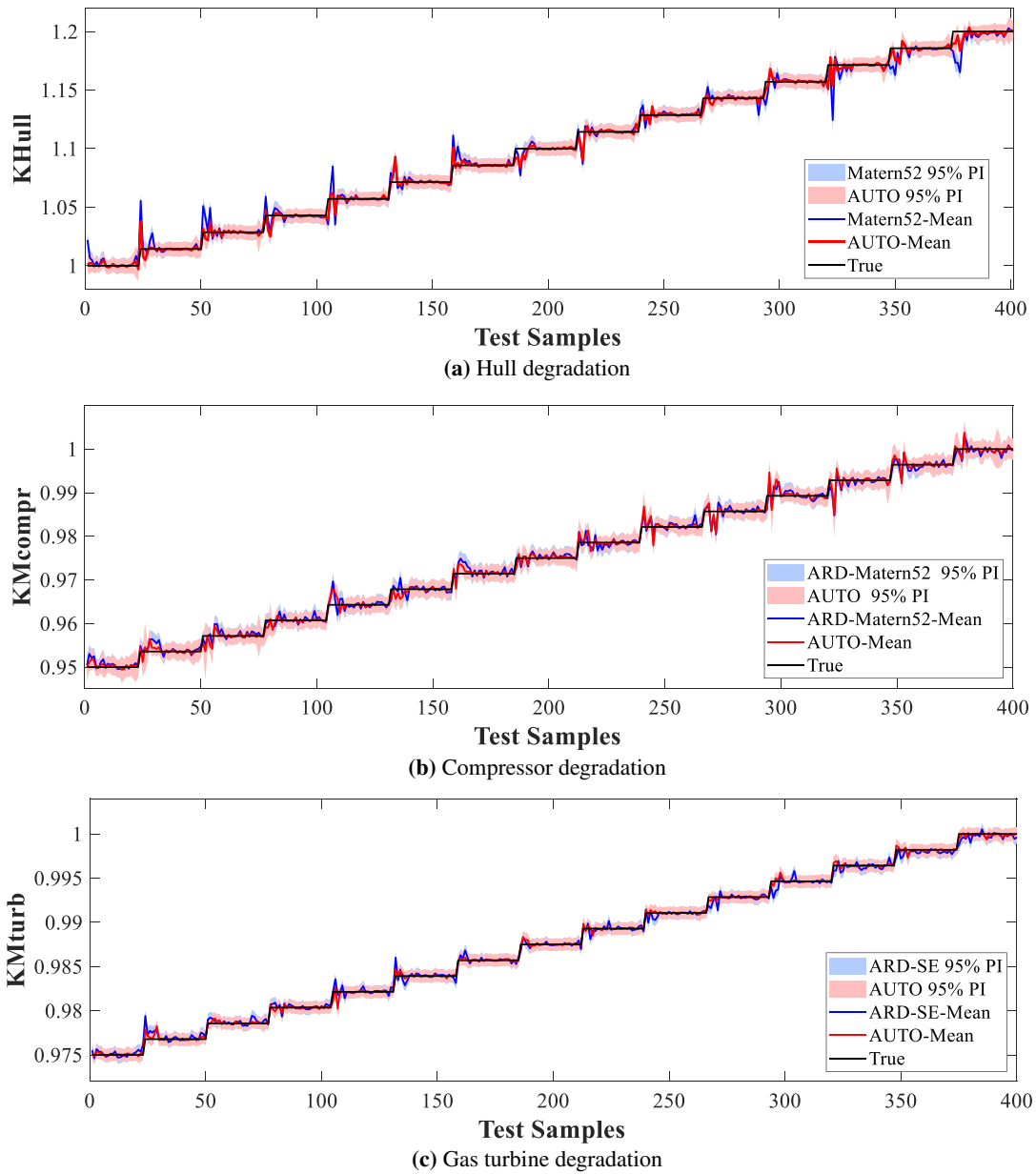


Figure 7: (Continued)

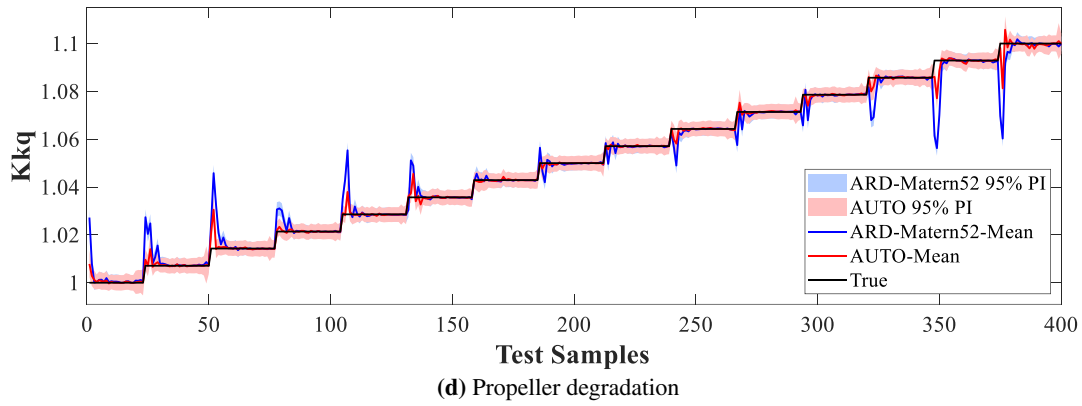


Figure 7: Prediction results for four degradation indicators, including predictive mean, 95% prediction intervals, and ground truth.

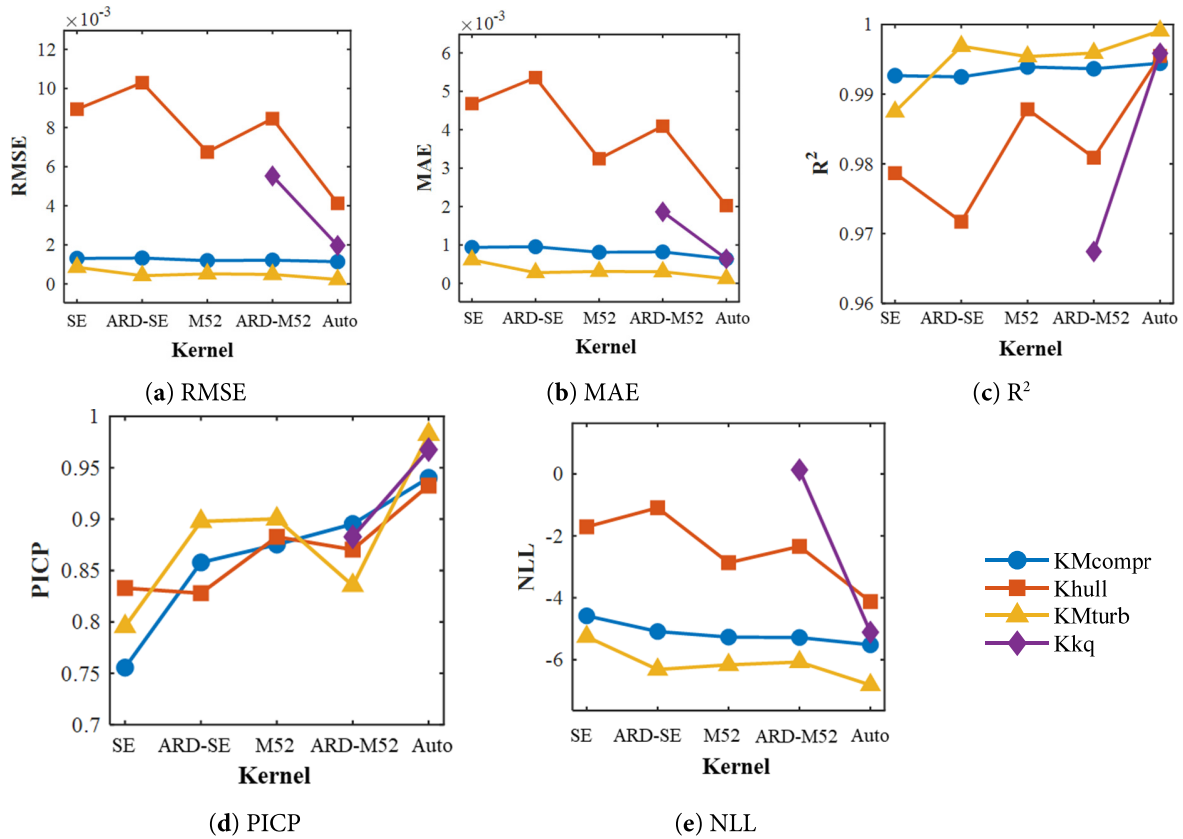


Figure 8: Quantitative evaluation of prediction performance across four degradation indicators in terms of RMSE, MAE, R^2 , PICP, and NLL under different kernel configurations. The kernel types include SE, ARD-SE, Matern52 (M52), ARD-Matern52 (ARD-52), and the automatically selected kernel (AUTO).

As shown in Fig. 7, the predictive mean closely follows the ground truth across all samples, indicating that the proposed method can capture the degradation trends of different ship propulsion system components. Compared with baseline kernels, the AUTO kernel provides stable predictions and well-calibrated prediction intervals.

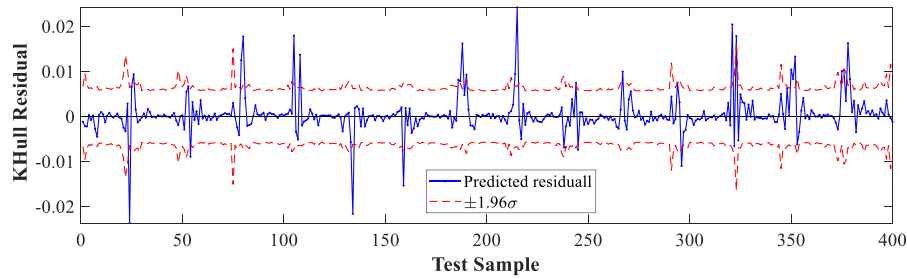
To further provide a clear quantitative evaluation, the prediction performance of the automatically selected kernel is summarized in Table 7. RMSE is used to evaluate point prediction accuracy, while PICP and NLL are used to assess uncertainty coverage and probabilistic calibration.

Table 7: Quantitative evaluation of prediction performance on the ship propulsion dataset.

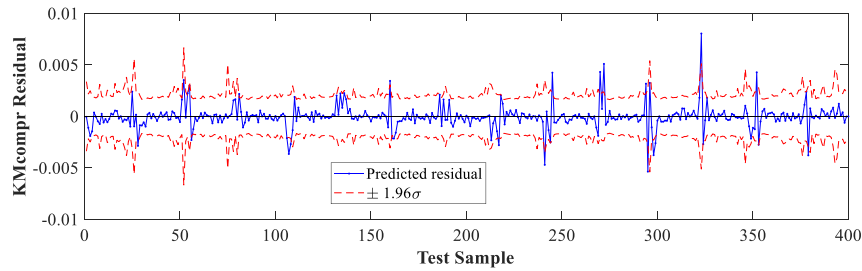
Degradation Indicator	RMSE	MAE	R ²	PICP	NLL
KMcompr	1.14×10^{-3}	0.64×10^{-3}	0.99444	0.94015	-5.5174
Khull	4.12×10^{-3}	2.03×10^{-3}	0.99547	0.93267	-4.1128
KMturb	0.23×10^{-3}	0.12×10^{-3}	0.99909	0.98254	-6.8164
Kkq	1.96×10^{-3}	1.62×10^{-3}	0.99587	0.96758	-5.1187

As shown in Table 7, the proposed method achieves low RMSE and MAE values for all four degradation indicators while maintaining high R² values, indicating accurate prediction performance. The PICP values are close to or higher than the nominal confidence level, and the NLL values remain low, showing that the proposed method can provide reliable uncertainty quantification in addition to point prediction.

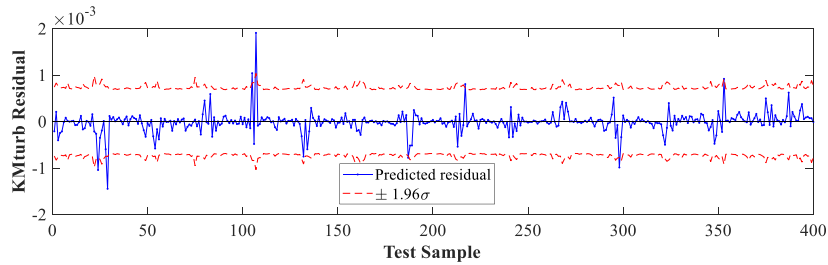
To further examine the prediction error characteristics, residual analysis is conducted, as shown in Fig. 9.



(a) Residuals of hull degradation



(b) Residuals of compressor degradation



(c) Residuals of gas turbine degradation

Figure 9: (Continued)

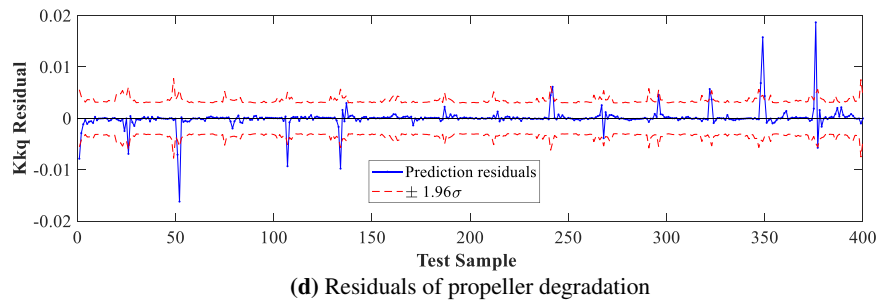


Figure 9: Residual analysis for four degradation indicators, including residual series and corresponding $\pm 1.96\sigma$ bounds.

As illustrated in Fig. 9, the residuals for all degradation indicators are generally centered around zero. Most residuals fall within the $\pm 1.96\sigma$ bounds, confirming that the uncertainty estimates are well calibrated.

4.4 Comparison with Related Ship Propulsion System Studies

To place the proposed method in the context of existing ship propulsion system health assessment studies, representative related works are summarized in Table 8. These studies are not directly comparable in a strict numerical sense because they differ in prediction targets, task formulations, and evaluation metrics. The comparison is therefore intended to clarify the methodological scope of the present work rather than to rank different methods by a single performance index.

Table 8: Comparison with related ship propulsion system studies.

Study	Research Object	Method	Target/Output	Uncertainty Quantification
Coraddu et al. [47]	Naval propulsion plant	RLS, SVR	GTC and GT degradation prediction	No
Cipollini et al. [46]	CODLAG naval propulsion system	Supervised DA + novelty detection	GT, GTC, HLL, and PRP degradation-state assessment	No
Tan et al. [49]	Naval propulsion plant	One-class SVM	Decay detection, decay degree estimation, and main decay direction identification	No
Berghout et al. [50]	CODLAG naval propulsion system	C-DBN based on online sequential ELM	GTC and GT degradation prediction	No
Present study	CODLAG naval propulsion system	Bayesian tree-structured GPR	Continuous degradation coefficient prediction of GTC, HLL, GT, and PRP	Yes

As shown in Table 8, previous studies have addressed ship propulsion system health assessment through deterministic degradation prediction, binary degradation-state assessment, or anomaly detection. In comparison, the proposed Bayesian tree-structured GPR method performs continuous degradation

coefficient prediction and further provides uncertainty quantification through prediction intervals, PICP, and NLL. This allows the prediction results to be assessed in terms of both accuracy and confidence reliability. Therefore, the proposed method complements existing studies by introducing an uncertainty-aware probabilistic framework for risk-informed health assessment and predictive maintenance of marine propulsion systems.

5 Conclusion

This study proposed a Bayesian tree-structured GPR method for automatic kernel selection and probabilistic degradation modeling of ship propulsion systems. Instead of using a fixed kernel, the proposed method constructs composite kernels from a predefined base-kernel library and selects suitable kernel structures using Bayesian evidence with validation-assisted refinement. For the PRONOSTIA bearing dataset, the automatically selected composite kernel achieved an RMSE of 0.092 and a PICP of 0.872. The NLL was also reduced to approximately -0.8 , indicating better uncertainty calibration than fixed-kernel GPR models. For the CODLAG ship propulsion simulation dataset, the RMSE values were 1.14×10^{-3} , 4.12×10^{-3} , 0.23×10^{-3} , and 1.96×10^{-3} for compressor, hull, gas turbine, and propeller degradation, respectively. The corresponding PICP values were 0.94015, 0.93267, 0.98254, and 0.96758. Taken together, these results indicate that the proposed method does not simply improve point prediction accuracy but also provides more reliable uncertainty estimates across different degradation scenarios. This ability to combine task-specific kernel adaptation with calibrated prediction intervals makes the method useful for early fault warning, health state assessment, and risk-aware predictive maintenance of ship propulsion systems.

Future research should further examine the proposed framework under online and full-scale operating conditions. In the present study, kernel search and model training are conducted in an offline manner, and a representative subset is used for the ship propulsion dataset to reduce the computational burden of repeated GPR training. A useful extension would be to develop an incremental or sparse version of the Bayesian kernel learning scheme, so that new sensor data can be incorporated without retraining the model from the beginning. This would make the method more suitable for real-time state estimation and dynamic RUL updating in onboard monitoring systems. In addition, the current implementation still relies on manually extracted features for bearing degradation modeling. Combining the proposed probabilistic kernel learning framework with deep feature extraction may improve its ability to handle high-dimensional vibration, acoustic, and thermal signals. Further validation using long-term operational data from real ship propulsion systems would also be necessary to assess the robustness of the method under changing loads, environmental disturbances, and incomplete sensor measurements.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the National Natural Science Foundation of China (Grant Nos. 52305255 and 52475252), Hebei Yanzhao Golden Platform Talent Gathering Project (Grant No. A2024007), Open Fund Project of Hubei Innovation Center for Digital Intelligence Engineering of Marine Equipment (Grant No. ZHJ250294), and the Young Elite Scientists Sponsorship Program by CAST. (Grant No. YESS20210467).

Author Contributions: Xinyu Jia: Methodology, Funding acquisition, Investigation, Conceptualization, Writing—original draft. Qingjie Wang: Methodology, Data curation, Writing—original draft, Investigation. Rongchuan Zhang: Validation, Writing—review & editing. Heng Ouyang: Writing—review & editing, Validation, Project administration. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Algorithm A1: Bayesian tree-structured kernel learning for GPR

Input:

Training dataset $\mathbf{D}_{\text{train}}$; validation set $\mathbf{D}_{\text{val}}$; Base kernel set $\mathbf{B} = \{k_{SE}, k_{LIN}, k_{PER}, k_{RQ}, k_{Matern}\}$
 Maximum search depth H_{max} ; coarse-screening width B_c ; Beam width B ; Complexity penalty parameter $\lambda_1, \lambda_2, \lambda_3$ Evidence tolerance threshold Δ

Output: Optimal kernel structure k^* and hyperparameters $\hat{\Theta}_k$

```

1: Initialize the kernel structure set and the evaluated structure set  $C_0 \leftarrow \mathbf{B}, \mathcal{E} \leftarrow \emptyset$ 
2:   For  $h = 1$  to  $H_{\text{max}}$ 
3:     Generate candidates by combining kernels in  $C_{h-1}$  using + and  $\times$ 
4:     Remove duplicate kernels and discard structures violating depth constraints
5:     for each candidate  $\kappa$  do
6:       Estimate hyperparameters via MAP optimization (Eq. (13))
7:       Compute BIC screening score  $S_{\text{BIC}}(\kappa)$  (Eq. (22))
8:     end for
9:     Retain the top  $B_c$  candidates ranked by  $S_{\text{BIC}}(\kappa)$ 
10:    for each retained candidate  $\kappa$  do
11:      Compute the Hessian at MAP estimate (Eq. (15))
12:      Approximate log evidence  $\log p(\mathbf{y} | \mathbf{X}, \kappa)$  via Laplace approximation (Eq. (17))
13:      Compute structural penalty  $\Omega(\kappa)$ 
14:      Compute Bayesian score  $S(\kappa) = \log p(\mathbf{y} | \mathbf{X}, \kappa) - \Omega(\kappa)$ 
15:      Add  $(\kappa, S(\kappa), \hat{\Theta}_\kappa)$  to  $\mathcal{E}$ 
16:    end for
17:    Select the top  $B$  candidates from the retained candidates by  $S(\kappa)$  as new beam  $C_h$ 
18:  End for
19:   $\kappa^* = \underset{\kappa \in \mathcal{E}}{\operatorname{argmax}} S(\kappa)$ 
20:  Define near-optimal set  $K_\Delta = \{\kappa \in \mathcal{E} : S(\kappa^*) - S(\kappa) \leq \Delta\}$ 
21:  if  $|K_\Delta| = 1$  then
22:     $\kappa_{\text{opt}} \leftarrow \kappa^*$ 
23:  else
24:    Compute the validation auxiliary score  $\text{AuxScore}(\kappa)$  for each  $\kappa \in K_\Delta$ 
25:     $\kappa_{\text{opt}} \leftarrow \underset{\kappa \in K_\Delta}{\operatorname{argmin}} \text{AuxScore}(\kappa)$ 
26:  end if
27:  Return  $\kappa_{\text{opt}}, \hat{\Theta}_{\kappa_{\text{opt}}}$ 

```

References

1. Li KX, Wang H, Yang Y, Li M. Resilience in maritime logistics: theoretical framework, research methodology, and indicator system. Ocean Coast Manag. 2024;259(1):107465. doi:10.1016/j.ocecoaman.2024.107465.

2. Velasco-Gallego C, Navas De Maya B, Matutano Molina C, Lazakis I, Cubo Mateo N. Recent advancements in data-driven methodologies for the fault diagnosis and prognosis of marine systems: a systematic review. *Ocean Eng.* 2023;284(1):115277. doi:10.1016/j.oceaneng.2023.115277.
3. Abdullah MU, Khan ZA. A multiscale overview of modelling rolling cyclic fatigue in bearing elements. *Materials.* 2022;15(17):5885. doi:10.3390/ma15175885.
4. Hou W, Peng Y. Adaptive ensemble Gaussian process regression-driven degradation prognosis with applications to bearing degradation. *Reliab Eng Syst Saf.* 2023;239(10):109479. doi:10.1016/j.ress.2023.109479.
5. Malhotra P, Ramakrishnan A, Anand G, Vig L, Agarwal P, Shroff G. LSTM-based encoder-decoder for multi-sensor anomaly detection. *arXiv:1607.00148.* 2016.
6. Zhang W, Li C, Peng G, Chen Y, Zhang Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech Syst Signal Process.* 2018;100(6):439–53. doi:10.1016/j.ymssp.2017.06.022.
7. Liang Q, Knutsen KE, Vanem E, Æsøy V, Zhang H. A review of maritime equipment prognostics health management from a classification society perspective. *Ocean Eng.* 2024;301(1):117619. doi:10.1016/j.oceaneng.2024.117619.
8. Kibrete F, Engida Woldemichael D, Shimels Gebremedhen H. Multi-Sensor data fusion in intelligent fault diagnosis of rotating machines: a comprehensive review. *Measurement.* 2024;232(4):114658. doi:10.1016/j.measurement.2024.114658.
9. Yin P, Han B, Wang J, Zhang Z. Bearing degradation state estimation based on EEMD improved alignment entropy and correlation analysis. In: *Proceedings of the 2024 Global Reliability and Prognostics and Health Management Conference (PHM-Beijing); 2024 Oct 11–13; Beijing, China.* p. 1–6.
10. Tai Y, Yan X, Jiang M, Zhang F, Geng X, Mu L. Data-driven method for predicting remaining useful life of bearings based on multi-layer perception neural network and bidirectional long short-term memory network. *Struct Durab Health Monit.* 2025;19(2):365–83. doi:10.32604/sdhm.2024.053998.
11. Velasco-Gallego C, Lazakis I. Mar-RUL: a remaining useful life prediction approach for fault prognostics of marine machinery. *Appl Ocean Res.* 2023;140(1):103735. doi:10.1016/j.apor.2023.103735.
12. Ren F, Du J, Chang D. Research on the bearing lifespan prediction method for ship propulsion shaft systems based on an enhanced domain adversarial neural network. *J Mar Sci Eng.* 2023;11(11):2128. doi:10.3390/jmse11112128.
13. Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital twin-driven product design, manufacturing and service with big data. *Int J Adv Manuf Technol.* 2018;94(9):3563–76. doi:10.1007/s00170-017-0233-1.
14. Hasan A, Asfihani T, Osen O, Bye RT. Leveraging digital twins for fault diagnosis in autonomous ships. *Ocean Eng.* 2024;292(10):116546. doi:10.1016/j.oceaneng.2023.116546.
15. Li C, Zou Z, Shen G, Chen M, Li D, Miao H, et al. A co-designed on-rotor sensing and tacholeless order tracking framework for robust fault diagnosis in marine electric propulsion systems. *Mech Syst Signal Process.* 2026;244:113751. doi:10.1016/j.ymssp.2025.113751.
16. Aizpurua JI, Knutsen KE, Heimdal M, Vanem E. Integrated machine learning and probabilistic degradation approach for vessel electric motor prognostics. *Ocean Eng.* 2023;275(3):114153. doi:10.1016/j.oceaneng.2023.114153.
17. Wu J, Hu K, Cheng Y, Wang J, Deng C, Wang Y. Ensemble recurrent neural network-based residual useful life prognostics of aircraft engines. *Struct Durab Heal Monit.* 2019;13(3):317–29. doi:10.32604/sdhm.2019.05571.
18. Dong X, Chen Q, Liu W, Wang D, Peng Z, Meng G. A systematic framework of constructing surrogate model for slider track peeling strength prediction. *Sci China Technol Sci.* 2024;67(10):3261–74. doi:10.1007/s11431-024-2764-5.
19. Lin YH, Yan PC, Zio E. Recent advances in uncertainty analysis for prognostics and remaining useful life prediction: a review. *Reliab Eng Syst Saf.* 2026;269(8):112110. doi:10.1016/j.ress.2025.112110.
20. Pan J, Sun B, Wu Z, Yi Z, Feng Q, Ren Y, et al. Probabilistic remaining useful life prediction without lifetime labels: a Bayesian deep learning and stochastic process fusion method. *Reliab Eng Syst Saf.* 2024;250:110313. doi:10.1016/j.ress.2024.110313.
21. Qian S, Shi Q, Yang C, Luo Z, Duan L, Zhao F. Damage identification method based on interval regularization theory. *Comput Meth Appl Mech Eng.* 2024;431(3):117288. doi:10.1016/j.cma.2024.117288.

22. Yang C, Ouyang H. A novel load-dependent sensor placement method for model updating based on time-dependent reliability optimization considering multi-source uncertainties. *Mech Syst Signal Process.* 2022;165(7):108386. doi:10.1016/j.ymssp.2021.108386.
23. Huang Y, Beck JL, Li H. Bayesian system identification based on hierarchical sparse Bayesian learning and Gibbs sampling with application to structural damage assessment. *Comput Meth Appl Mech Eng.* 2017;318(7):382–411. doi:10.1016/j.cma.2017.01.030.
24. Kosikova AM, Sedehi O, Papadimitriou C, Katafygiotis LS. Bayesian structural identification using Gaussian process discrepancy models. *Comput Meth Appl Mech Eng.* 2023;417:116357. doi:10.1016/j.cma.2023.116357.
25. Meng X, Beck JL, Huang Y, Li H. Adaptive meta-learning stochastic gradient Hamiltonian Monte Carlo simulation for Bayesian updating of structural dynamic models. *Comput Meth Appl Mech Eng.* 2025;437(7):117753. doi:10.1016/j.cma.2025.117753.
26. BahooToroody A, Abaei MM, Banda OV, Kujala P, De Carlo F, Abbassi R. Prognostic health management of repairable ship systems through different autonomy degree; From current condition to fully autonomous ship. *Reliab Eng Syst Saf.* 2022;221(2):108355. doi:10.1016/j.res.2022.108355.
27. BahooToroody A, Abaei MM, Zio E, Goerlandt F, Chaal M. Gaussian process latent variable model and Bayesian inference for non-parametric failure modeling applied to ship engine. *Reliab Eng Syst Saf.* 2026;265(1):111611. doi:10.1016/j.res.2025.111611.
28. Niu T, Xu Z, Luo H, Zhou Z. Hybrid Gaussian process regression with temporal feature extraction for partially interpretable remaining useful life interval prediction in Aeroengine prognostics. *Sci Rep.* 2025;15(1):11057. doi:10.1038/s41598-025-88703-z.
29. Rasmussen CE, Williams CKI. *Gaussian processes for machine learning.* Cambridge, MA, USA: The MIT Press; 2005.
30. Duvenaud D, Lloyd JR, Grosse R, Tenenbaum JB, Ghahramani Z. Structure discovery in nonparametric regression through compositional kernel search. In: *Proceedings of the 30th International Conference on Machine Learning*; 2013 Jun 16–21; Atlanta, GA, USA. p. 1166–74.
31. Kuok SC, Yao SA, Yuen KV, Ding Z, Duan Y. Gaussian process with Bayesian kernel selection for seismic attenuation modeling. *ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng.* 2025;11(4):04025054. doi:10.1061/ajrua6.rueng-1579.
32. Li Y, Wan D, Wang R, Ni B, Hu D, Jiang C. A self-adaptive Gaussian process regression approach for temperature fluctuation predictions in T-junction pipes of sodium-cooled fast reactors. *Sci China Technol Sci.* 2026;69(1):1120102. doi:10.1007/s11431-025-3077-y.
33. Sun S, Zhang G, Wang C, Zeng W, Li J, Grosse R. Differentiable compositional kernel learning for Gaussian processes. In: *Proceedings of the 35th International Conference on Machine Learning*; 2018 Jul 10–15; Stockholm, Sweden. p. 4828–37.
34. Wilson AG, Adams RP. Gaussian process kernels for pattern discovery and extrapolation. In: *Proceedings of the 30th International Conference on Machine Learning*; 2013 Jun 16–21; Atlanta, GA, USA. p. 1067–75.
35. Jia X, Papadimitriou C. Hierarchical Bayesian learning framework for multi-level modeling using multi-level data. *Mech Syst Signal Process.* 2022;179(3):109179. doi:10.1016/j.ymssp.2022.109179.
36. Kuok SC, Yao SA, Yuen KV, Yan WJ, Girolami M. Bayesian generative kernel Gaussian process regression. *Mech Syst Signal Process.* 2025;227(1):112395. doi:10.1016/j.ymssp.2025.112395.
37. Jin SS. Compositional kernel learning using tree-based genetic programming for Gaussian process regression. *Struct Multidiscip Optim.* 2020;62(3):1313–51. doi:10.1007/s00158-020-02559-7.
38. He F, Long X, Liu J, Jiang C. A deep convolutional framework for unpaired multi-fidelity fusion through structured panel encoding. *Struct Multidiscip Optim.* 2026;69(4):81. doi:10.1007/s00158-026-04293-y.
39. Smith CAB, Box GEP, Tiao GC. Bayesian inference in statistical analysis. *Int Stat Rev.* 1975;43(2):242. doi:10.2307/1402913.
40. Jia X, Papadimitriou C. Data features-based Bayesian learning for time-domain model updating and robust predictions in structural dynamics. *Mech Syst Signal Process.* 2025;224:112197. doi:10.1016/j.ymssp.2024.112197.

41. Jia X, Wang Q, Li X, Han X. A two-stage data-driven approach for remaining useful life prediction. *Acta Mech Sin.* 2026;42(8):725810. doi:10.1007/s10409-025-25810-x.
42. Papadimitriou C, Beck JL, Katafygiotis LS. Asymptotic expansions for reliability and moments of uncertain systems. *J Eng Mech.* 1997;123(12):1219–29. doi:10.1061/(asce)0733-9399(1997)123:12(1219).
43. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising model component complexity: a principled, practical approach to constructing priors. *Statist Sci.* 2017;32(1):1–28. doi:10.1214/16-sts576.
44. Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Chebel-Morello B, Zerhouni N, et al. PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: *Proceedings of the IEEE International Conference on Prognostics and Health Management, PHM'12*; 2012 Jun 18–21; Denver, CO, USA.
45. Sun B, Hu W, Wang H, Wang L, Deng C. Remaining useful life prediction of rolling bearings based on CBAM-CNN-LSTM. *Sensors.* 2025;25(2):554. doi:10.3390/s25020554.
46. Cipollini F, Oneto L, Coraddu A, Murphy AJ, Anguita D. Condition-based maintenance of naval propulsion systems: data analysis with minimal feedback. *Reliab Eng Syst Saf.* 2018;177(1):12–23. doi:10.1016/j.res.2018.04.015.
47. Coraddu A, Oneto L, Ghio A, Savio S, Anguita D, Figari M. Machine learning approaches for improving condition-based maintenance of naval propulsion plants. *Proc Inst Mech Eng Part M J Eng Marit Environ.* 2016;230(1):136–53. doi:10.1177/1475090214540874.
48. Coraddu A, Oneto L, Ghio A, Savio S, Anguita D, Figari M. Condition based maintenance of naval propulsion plants. *UCI Mach Learn Repository.* 2014. doi:10.24432/C5K31K.
49. Tan Y, Niu C, Tian H, Hou L, Zhang J. A one-class SVM based approach for condition-based maintenance of a naval propulsion plant with limited labeled data. *Ocean Eng.* 2019;193(1):106592. doi:10.1016/j.oceaneng.2019.106592.
50. Berghout T, Mouss LH, Bentrchia T, Elbouchikhi E, Benbouzid M. A deep supervised learning approach for condition-based maintenance of naval propulsion systems. *Ocean Eng.* 2021;221(7):108525. doi:10.1016/j.oceaneng.2020.108525.