Remaining Useful Life Prediction Using Constraint-Guided Learning with Limited Physical Knowledge

Ilyas Lemmens[1,2], Wout Rombouts[1,2], Quinten Van Baelen[1,2], Peter Karsmakers[1,2], and Mathias Verbeke[1,2]

[1] *KU Leuven, Department of Computer Science, DTAI - Leuven.AI, Belgium*
firstname.lastname@kuleuven.be

[2] *Flanders Make@KU Leuven - Leuven, Belgium*

ABSTRACT

Unexpected failures in industrial assets can cause significant downtime and costs. Anticipating the Remaining Useful Life (RUL) enables proactive maintenance strategies that mitigate such risks. While deep learning models have proven themselves to be adept at RUL prediction, they do not inherently guarantee physically consistent predictions. This paper explores two alternative approaches for incorporating physical constraints into data-driven RUL prediction models. The first approach extends prior work on physics-guided loss functions by integrating domain knowledge directly into the training objective. Physical assumptions are encoded as additional penalty terms that act as regularizers, discouraging physically implausible behavior while allowing trade-offs with predictive accuracy. The second approach leverages Constraint-Guided Gradient Descent (CGGD), a recent optimization framework which enforces constraints at the optimization level rather than through the loss function. CGGD monitors constraint satisfaction during training and dynamically modifies gradient updates only when violations occur, steering the solution back into the feasible region. Both methods aim to improve model interpretability and robustness without requiring detailed or fully specified physical knowledge, making them applicable to a wide range of industrial settings. We evaluate these strategies across multiple experimental setups, comparing standard predictive accuracy with additional physics-based evaluation metrics that assess adherence to physical assumptions. The findings provide useful insights into the advantages and limitations of constraint enforcement techniques and their overall impact on predictions, contributing to the development of trustworthy Prognostics and Health Management (PHM) models.

1. INTRODUCTION

Remaining Useful Life (RUL) prediction is a central challenge in the Prognostics and Health Management (PHM) field, enabling maintenance strategies that minimize downtime and operational costs (Biggio & Kastanis, 2020). Bearings, in particular, are a key focus due to their susceptibility to early failure. Bearing degradation can severely affect load support and friction reduction, ultimately shortening the overall lifespan of the machine (Chennana et al., 2025). Data-driven Machine Learning (ML) models have shown strong performance in this setting by learning degradation patterns directly from sensor data (Wang, Yan, Du, Li, & Chen, 2025). Recent architectures continue to improve by capturing increasingly complex temporal dependencies (Wu, Wu, Tan, & Xu, 2024; Song et al., 2024). However, despite their predictive power, such models do not inherently adhere to fundamental physical laws. This can result in unrealistic behavior, such as non-monotonic predictions or inconsistencies with degradation indicators, especially in noisy or data-limited environments. Furthermore, these models are often viewed as black boxes, offering limited transparency into the reasoning behind their decisions (Herwig, Borghesani, Smith, & Peng, 2025; Youness & Aalah, 2023).

To address these challenges, recent research seeks to integrate physical knowledge into data-driven prognostics to increase both reliability and explainability (Raissi, Perdikaris, & Karniadakis, 2019). One promising direction is to impose domain knowledge in the form of constraints on the ML model's behavior (Park, Kim, & Choi, 2024). Constraints aim to limit ML models to predictions that are valid, while still leveraging the flexibility of deep learning.

In this work, we investigate how different constraint integration strategies affect data-driven RUL prediction. We focus on two known physical assumptions, monotonic degradation and consistency with vibration energy, studying how they influence predictions when implemented through (i) physics-guided loss functions and (ii) Constraint-Guided Gradient Descent

(CGGD). Experiments are conducted on two bearing degradation datasets, and performance is evaluated using traditional error-based metrics as well as physics-oriented measures of trendability and robustness (Tefera, Van Baelen, Meire, Luca, & Karsmakers, 2025).

The main contributions of this work are:

- A comparison of two constraint-integration strategies, physics-guided loss functions and Constraint-Guided Gradient Descent, evaluated under identical model and data conditions.
- An extended study across two bearing datasets with different degradation behavior, demonstrating how constraint effectiveness depends on the underlying physical characteristics of the data.
- An evaluation using both accuracy-based and physics-based metrics, providing a more comprehensive assessment of constraint mechanisms.

The remainder of this paper proceeds as follows. Section 2 situates this study within existing work on physics-inspired and constraint-based RUL modeling. Section 3 introduces the two constraint formulations and how they implement the monotonicity and energy assumptions. Section 4 presents the experimental setup. This section includes the datasets, feature extraction process, model architecture, and evaluation metrics. Section 5 reports the results and analyzes how each constraint influences RUL predictions under different degradation characteristics. Section 6 concludes with the main findings and future directions.

2. RELATED WORK

Recent work in PHM increasingly integrates physical knowledge to improve reliability, interpretability, and generalization. In this context, several strategies have been proposed for introducing prior knowledge into different levels of a neural network (Wang et al., 2025; Li, Zhang, Li, & Si, 2024). First, *data-level integration* enriches data through simulation models or carefully engineered features (Y. Chen, Rao, Feng, & Zuo, 2022; H. Lu et al., 2023). Second, *architecture-level integration* embeds physics directly into the model through physically motivated layers or constrained neurons (Yucesan & Viana, 2022; J. Chen & Liu, 2021).

Third, *loss-level integration* incorporates physics by manipulating the training objective. This approach can be divided into two subcategories, depending on how physics is imposed in the loss function: Physics-Informed Neural Networks (PINNs) implement explicit governing equations within the loss function. These methods require detailed machine and/or physics knowledge (Raissi et al., 2019; Liao, Chen, Wen, & Zhao, 2023). In contrast, physics-guided losses incorporate flexible, high-level constraints that do not rely on full physical models and typically do not require detailed knowledge (Zhu et al., 2024; He, Su, Zio, Fan, & Zhang, 2022; W. Lu, Wang, Zhang, & Gu, 2024).

Finally, *Constraint-guided optimization* has emerged as an alternative means of enforcing physics-guided losses. Rather than embedding the constraints directly into the loss, CGGD enforces them at the optimization level whenever violations occur (Van Baelen & Karsmakers, 2023; Rombouts, Van Baelen, & Karsmakers, 2025). This avoids loss-weight tuning and enables the use of non-differentiable constraints.

3. METHODOLOGY

In this work, we consider two commonly used physical assumptions for bearing degradation under normal operating conditions: (i) RUL is expected to decrease monotonically over time, and (ii) changes in RUL are expected to be consistent with changes in vibration energy. These constraints can be enforced either via penalty terms in the loss function (Section 3.1) or via a constrained optimization solver (Section 3.2), enabling a direct comparison between the two constraint integration strategies.

3.1. Physics-guided loss functions

To incorporate the physical assumptions mentioned above into a supervised learning setting, physics-guided loss functions implement them as additional penalty terms in the training objective. The total loss is defined as

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \alpha\, \mathcal{L}_{physics}, \tag{1}$$

where α controls the influence of the constraint term relative to the supervised RUL error $\mathcal{L}_{MSE}$. The term $\mathcal{L}_{physics}$ penalizes violations of the monotonicity and energy consistency assumptions.

Monotonicity constraint. The monotonicity assumption is implemented by penalizing increases in predicted RUL between consecutive time steps (W. Lu et al., 2024). This local formulation is sufficient to guarantee global monotonic behavior, as enforcing monotonicity between all adjacent time steps implies a monotonically decreasing trajectory:

$$\mathcal{L}_{physics} = \frac{1}{N-1} \sum_{i=1}^{N-1} \max\left(0,\, \hat{y}_{i+1} - \hat{y}_i\right), \tag{2}$$

where N is the number of samples in a run and $\hat{y}_i$ the predicted RUL at time i.

Energy-based constraint. To implement consistency between RUL and vibration energy, we relate changes in RUL to changes in the Root Mean Square (RMS) of the accelerometer signal. Increases in vibration energy are expected to coincide with decreases in RUL and vice versa. Following previous energy-based formulations (Zhu et al., 2024), we penalize

cases where both quantities change in the same direction:

$$\mathcal{L}_{\text{physics}} = \frac{1}{N-1}\sum_{i=1}^{N-1}\max\Big(0, (\hat{y}_i - \hat{y}_{i+1})(\text{RMS}_i - \text{RMS}_{i+1})\Big). \tag{3}$$

where N is the number of samples, $\hat{y}_i$ is the predicted RUL at time i, and RMS_i is the RMS value at time i.

3.2. Constraint-Guided Gradient Descent

While physics-guided losses enforce constraints through additional penalty terms, CGGD incorporates the same physical assumptions at the level of the optimization dynamics. Instead of modifying the loss function, CGGD adjusts the parameter update whenever a constraint is violated. This framework, introduced in Van Baelen and Karsmakers (2023) and recently applied to Health Indicator learning (Tefera et al., 2025), enables constraint-consistent learning without requiring differentiable constraints or loss-weight tuning.

Given the model gradient

$$g = \nabla_\theta \mathcal{L}_{model}, \tag{4}$$

CGGD evaluates the constraint $\mathcal{C}$. If the current model yields a prediction within the feasible region, the update proceeds unchanged compared to the update obtained from the gradient of $\mathcal{L}_{MSE}$. Otherwise, a constraint direction d_C with unit norm is computed and the update direction is modified to

$$\Delta\theta = g + R\, d_C\, \max(\|g\|, \epsilon)\,, \tag{5}$$

where $R > 1$ is a fixed constant (final solution is not sensitive to a specific value) that lets the constraint direction dominate the update to make sure the updated model comes closer to the feasible region, $\epsilon > 0$ prevents the update from vanishing when $\|g\|$ is small, and $d_C = 0$ whenever the constraint is satisfied.

The CGGD framework allows multiple constraint types to be incorporated in this way, including the monotonicity and energy-consistency assumptions introduced earlier. Each constraint independently checks whether its feasibility condition is violated for the current batch and contributes a correction d_C in the appropriate direction to the update. This mechanism ensures that constraint satisfaction is prioritized without modifying the underlying loss formulation. For the monotonicity constraint, the direction is determined by comparing the rank of each predicted RUL value to its expected rank in time: if a later sample is assigned a higher RUL than an earlier one, the update adjusts their ordering to restore a consistent degradation pattern. For the energy-consistency constraint, the same rank-based mechanism is applied using RMS as the reference, ensuring that predictions remain consistent with the expected inverse relationship between vibration energy and RUL.

4. EXPERIMENTAL SETUP

4.1. Datasets

This study evaluates the proposed RUL prediction methods on two differing bearing degradation datasets: the public benchmark dataset PHM2012 (Nectoux et al., 2012) and a private dataset collected by Flanders Make.

4.1.1. PHM2012 dataset

The PHM2012 dataset was collected on the PRONOSTIA test rig developed by FEMTO-ST for bearing RUL studies. Bearings were subjected to accelerated degradation under controlled radial loads while vibration signals were recorded via accelerometers on the bearing housing. Measurements were taken every 10s at 25.6 kHz, with 2560 samples per acquisition. The dataset includes multiple bearings under three operating conditions; we focus on Condition 1 and use Bearings 1-1 and 1-2 for training, Bearing 1-7 for validation, and Bearings 1-3 and 1-4 for testing. Fig. 1 shows the PRONOSTIA setup and sensor placement.

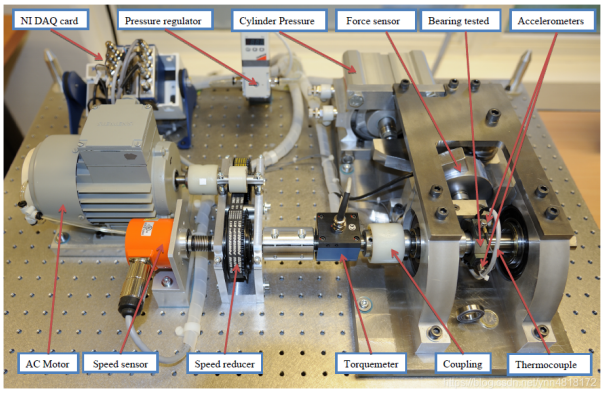


Figure 1. PHM2012 Bearing setup (Nectoux et al., 2012)

4.1.2. Private dataset

The dataset contains high-frequency acceleration measurements collected during accelerated lifetime tests of rolling bearings. The test rigs applied a radial load of 9 kN while the shaft rotated at 2000 rpm. Accelerometers were attached to the bearing housing, and signals were sampled at 50 kHz. Tests were stopped when peak vibrations exceeded 20g. Each of the 7 test rigs was used to perform 10 runs: 7 faulty bearings with a small pre-existing outer race indent and 3 healthy bearings. Faulty bearings were run until the stopping condition, while healthy bearings were run for 2 hours.

For the experiments in this study, we focused on the first test rig setup as it contained the most accurate and complete sensor

data. The available faulty runs were split into 2 training, 1 validation, and 2 test subsets consistent with the PHM2012 dataset. Since no exact RUL ground truth was provided, a kurtosis-based thresholding method was applied to segment the healthy and damaged regions of each run (W. Lu et al., 2024). Fig. 2 shows the test rig built by Flanders Make and sensor placement.

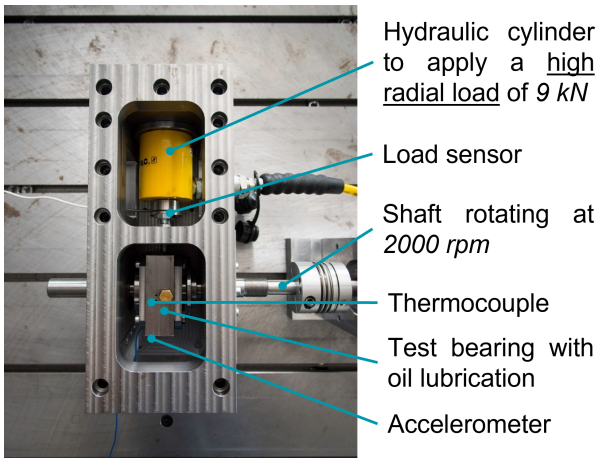


Figure 2. Private dataset bearing setup (Meire et al., 2022)

4.2. Pre-processing and model architecture

For each dataset, vibration signals were pre-processed, windowed and transformed into statistical features to serve as model inputs. A total of 32 features were extracted per window, mainly those found in Zhu et al. (2024) and Zhang, Zhang, and Xu (2016) : 18 time-domain, 6 frequency-domain, and 8 time-frequency features. The time and frequency-domain features are summarized in Table 1. The time-frequency domain features represent the signal energy within eight distinct wavelet decomposition bands. To reduce input dimensionality, a Random Forest regressor was used to rank feature importance with respect to RUL, retaining the 16 most informative features for training.

The predictive model is a two-layer Bidirectional Long Short-Term Memory (BiLSTM) architecture (Graves & Schmidhuber, 2005) with 128 hidden units per layer, followed by a fully connected output layer producing a single RUL estimate. The BiLSTM architecture was chosen for its ability to model temporal dependencies in both forward and backward directions, improving generalization in bearing degradation prediction (Luo & Zhang, 2022; Zhu et al., 2024).

4.3. Evaluation metrics

Predictive accuracy is evaluated using the Root Mean Squared Error (RMSE), which is the most commonly reported metric in RUL prediction literature (Song et al., 2024). RMSE quantifies the average magnitude of prediction errors and enables direct comparison with prior studies. It is defined as

Table 1. Time and frequency domain features

Time-Domain Features			
Feature	**Formula**	**Feature**	**Formula**
Mean	$\bar{x}$	Max	$x_{\max}$
Min	$x_{\min}$	Absolute Mean	$\overline{\lvert x\rvert}$
Smr	$\left(\overline{\sqrt{\lvert x\rvert}}\right)^2$	Variance	σ^2
Standard Deviation	σ	Crest Factor	$x_{\max}/\text{rms}$
Latitude Factor	$x_{\max}/\text{smr}$	Impulse Factor	$x_{\max}/\overline{\lvert x\rvert}$
Skewness	$\frac{\overline{(x-\bar{x})^3}}{\sigma^3}$	Kurtosis	$\frac{\overline{(x-\bar{x})^4}}{\sigma^4}$
Peak-to-Peak	$x_{\max} - x_{\min}$	Steepness	$\overline{(x-\bar{x})^4}/\sigma$
Pulse Index	$x_{\max}/\overline{\lvert x\rvert}$	Peak Index	$x_{\max}/\text{rms}$
Margin Index	$x_{\max}/\text{rms}^2$	Waveform Index	$\text{rms}/\overline{\lvert x\rvert}$
Frequency-Domain Features			
Feature	**Formula**	**Feature**	**Formula**
Frequency Center	$f = \frac{\sum f_i P_i}{\sum P_i}$	Spectral Deviation	$\sigma_f = \sqrt{\frac{\sum (f_i - f)^2 P_i}{\sum P_i}}$
Spectral Skewness	$\frac{\sum (f_i - f)^3 P_i}{(\sigma_f)^3 \sum P_i}$	Spectral Kurtosis	$\frac{\sum (f_i - f)^4 P_i}{(\sigma_f)^4 \sum P_i}$
Spectral RMS	$\sqrt{\frac{\sum f_i^2 P_i}{\sum P_i}}$	Frequency Mean	$\bar{P}$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}, \tag{6}$$

where $\hat{y}_i$ denotes the predicted RUL and y_i the corresponding ground-truth RUL at time step i, and N is the total number of time steps. RMSE implicitly assumes a smooth degradation process as ground truth, penalizing large deviations quadratically.

To assess the physical consistency of the models, especially regarding the effect of physics-guided losses, we quantify alignment using two dedicated metrics. Tefera et al. (2025) originally applied these metrics to Health Indicators, but they are equally relevant for RUL predictions.

Trendability. Measures how consistently the predicted RUL decreases as the system degrades. It is computed using the Spearman rank correlation coefficient (Carino, Zurita, Delgado, Ortega, & Romero-Troncoso, 2015):

$$\text{Trendability} = 1 - \frac{6\sum_{i=1}^{N}\left(\text{rank}(\hat{y}_i) - \text{rank}(i)\right)^2}{N(N^2-1)}, \tag{7}$$

where $\text{rank}(\hat{y}_i)$ is the rank of the predicted RUL at time step i, and $\text{rank}(i)$ is the rank of the corresponding time index. A value of -1 indicates a perfectly decreasing trajectory, thus more negative values correspond to higher trendability.

Robustness. Measures the stability and smoothness of the predicted RUL trajectory by comparing it to a smoothed version of itself, denoted $\tilde{y}_i$ using LOESS (Cleveland & Devlin, 1988):

$$\text{Robustness} = \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\frac{|\hat{y}_i - \tilde{y}_i|}{\hat{y}_i}\right). \tag{8}$$

Values range from 0 to 1, with higher values indicating smoother predictions.

4.4. Constraints and training

Three variants of the same BiLSTM architecture were trained to evaluate the proposed methods: (i) a baseline model trained with only the MSE loss, (ii) physics-guided models trained using either the monotonic or energy-based loss functions described in Section 3.1, and (iii) models trained with CGGD as the optimizer as described in Section 3.2.

For the baseline BiLSTM, standard MSE loss was applied without any additional constraints. Physics-guided models were trained using the monotonic and energy-based losses with weight parameters selected to balance the influence of the physics-guided term relative to the MSE loss. The chosen weights were selected via a coarse grid search to ensure notable influence on the predictions. For PHM2012, the weights were $\alpha \in \{1, 5, 15\}$ for both monotonic and energy-based losses. For the private dataset, larger weights had to be applied to obtain meaningful effects: monotonic loss used $\alpha \in \{5, 20, 50\}$ and energy-based loss used $\alpha \in \{100, 1000, 10000\}$. Both baseline and physics-guided models were trained using the Adam optimizer with learning rate 10^{-4} and early stopping based on the validation set.

CGGD models, implemented in this study via the CONGRADS[1] toolbox, do not require loss weighting. The following constraints were applied during training:

1. **RUL boundaries:**

$$-0.05 \leq \hat{y}_i \leq 1.05. \tag{9}$$

2. **Starting boundary constraint:**

$$\hat{y}_i \geq 0.9 \qquad \forall\, i \in \text{first } 5\% \text{ of the trajectory.} \tag{10}$$

3. **Failure boundary constraint:**

$$\hat{y}_i \leq 0.1 \qquad \forall\, i \in \text{last } 5\% \text{ of the trajectory.} \tag{11}$$

4. **Degradation constraints:**

$$(\hat{y}_i - \hat{y}_{i+1})(\mathrm{RMS}_i - \mathrm{RMS}_{i+1}) \leq 0 \quad \text{(Energy)}, \tag{12}$$

or

$$\hat{y}_{i+1} \leq \hat{y}_i \qquad \text{(Monotonicity)}. \tag{13}$$

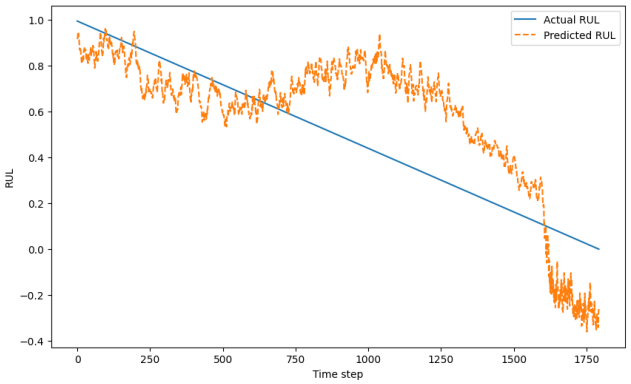
For CGGD models, training was performed with "ZeroLoss", meaning that the MSE penalty was set to zero and the model was optimized solely with respect to the constraints averaged over each batch. This allows the influence of the constraints on predictive behavior to be isolated. Initial tests with MSE led to predictions similar to the baseline models but they never improved accuracy, slowed convergence, and reduced adherence to the physical constraints. ZeroLoss generally enforces stricter satisfaction of the constraints while maintaining competitive evaluation metrics. Although CGGD-ZeroLoss resembles unsupervised degradation modelling, the combined constraints encourage a decreasing RUL trajectory from 1 to 0 which does not necessarily align with the usual linear decreasing target.

[1] https://congrads.readthedocs.io/en/latest/index.html

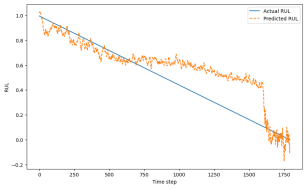
5. RESULTS AND DISCUSSION

5.1. PHM2012

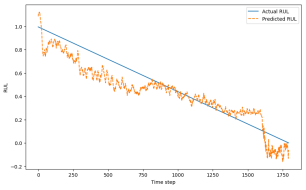
Table 2 reports the RMSE, trendability, and robustness metrics for the baseline BiLSTM and the physics-guided variants. All results are averaged over ten random seeds in order to reduce the influence of training variability. The quantitative trends are further illustrated by the prediction trajectories shown in Fig. 3.



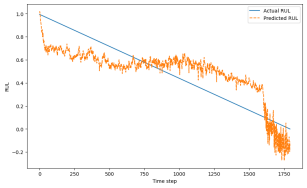
(a) Baseline



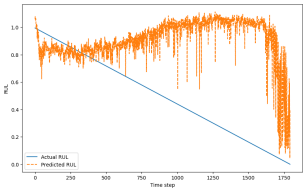
(b) Mono 5.0



(c) Mono CGGD



(d) Energy 5.0



(e) Energy CGGD

Figure 3. Effect of monotonic and energy physical loss functions on predictions for Bearing1-3 from the PHM2012 dataset. RUL values are normalized between 0 and 1. One time step is equal to 10 seconds.

Table 2. Performance metrics per model for the PHM2012 dataset. All bold values indicate the best performance per metric and are statistically significant compared to the baseline using a paired statistical test ($p < 0.05$).

Model α	**Bearing1_3**			**Bearing1_4**		
	RMSE	Trendability	Robustness	RMSE	Trendability	Robustness
Baseline	0.165 ± 0.027	-0.856 ± 0.063	0.773 ± 0.013	0.226 ± 0.018	-0.807 ± 0.052	0.744 ± 0.018
Mono 1.0	0.182 ± 0.058	-0.841 ± 0.107	0.777 ± 0.019	0.215 ± 0.022	-0.838 ± 0.014	0.776 ± 0.024
Mono 5.0	0.183 ± 0.030	-0.941 ± 0.028	0.815 ± 0.014	0.209 ± 0.015	-0.816 ± 0.013	0.806 ± 0.012
Mono 15.0	0.217 ± 0.007	-0.958 ± 0.032	0.879 ± 0.013	0.219 ± 0.004	-0.834 ± 0.011	0.886 ± 0.006
Mono CGGD	$\mathbf{0.125 \pm 0.038}$	$\mathbf{-0.972 \pm 0.014}$	0.812 ± 0.032	$\mathbf{0.180 \pm 0.031}$	$\mathbf{-0.868 \pm 0.014}$	0.663 ± 0.083
Energy 1.0	0.167 ± 0.025	-0.827 ± 0.073	0.779 ± 0.012	0.211 ± 0.017	-0.789 ± 0.052	0.744 ± 0.014
Energy 5.0	0.174 ± 0.018	-0.868 ± 0.082	0.804 ± 0.009	0.200 ± 0.011	-0.741 ± 0.041	0.791 ± 0.007
Energy 15.0	0.225 ± 0.013	-0.685 ± 0.167	0.808 ± 0.007	0.247 ± 0.006	-0.574 ± 0.053	0.843 ± 0.008
Energy CGGD	0.526 ± 0.024	0.467 ± 0.027	$\mathbf{0.893 \pm 0.007}$	0.509 ± 0.029	0.377 ± 0.359	$\mathbf{0.917 \pm 0.006}$

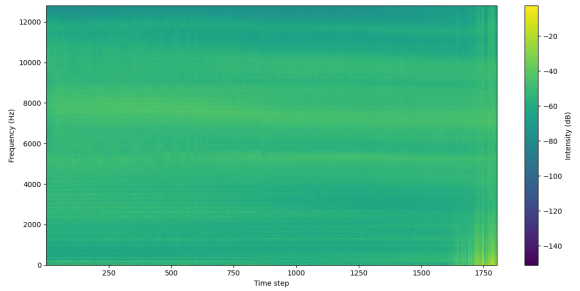


Figure 4. Spectrogram of the raw accelerometer signal for Bearing1-3 in PHM2012. One time is equal to 10 seconds.

For the monotonic loss, the results reveal a clear trade-off between predictive accuracy and physical consistency. As the weight α increases, both trendability and robustness improve consistently across the two test bearings. This confirms that the constraint effectively suppresses upward deviations and local fluctuations in the RUL trajectory. A visual comparison between the baseline predictions in Fig. 3a and the constrained model in Fig. 3b further supports this observation: the monotonic loss removes the physically impossible "arc" present in the baseline model. However, the effect on the RMSE metric is case-dependent. On Bearing1-4, low to moderate values of α lead to improved accuracy relative to the baseline. In contrast, on Bearing1-3, RMSE increases steadily with larger α. These observations reveal the importance of carefully balancing α: while stronger monotonic enforcement enhances physical plausibility, overly large weights can compromise predictive accuracy, particularly when the baseline model already provides a good fit to the degradation trend.

The monotonic CGGD avoids this trade-off. As shown in Table 2 and Fig. 3c, Monotonic CGGD achieves the lowest RMSE and highest trendability on both bearings, despite requiring neither loss-weight tuning nor an explicit MSE objective. In the CGGD-ZeroLoss setting, updates occur only when monotonicity or boundary constraints are violated. The resulting trajectory therefore reflects how the PHM2012 features evolve under constraint enforcement rather than continuous penalization. In this dataset, the extracted features are fairly stable for most of the run and experience a relatively sharp change near end of life. This behavior is also clearly visible in the vibration spectrogram in Fig. 4, which shows that the frequency content remains nearly constant until a broadband increase appears near time step 1750, right before failure. CGGD appears able to enforce monotonicity over the smaller deviations in early and mid-life data while respecting the enforced start- and end-of-life bounds, guiding the prediction towards a smooth decreasing trajectory.

Finally, both energy-based approaches tend to increase robustness but have limited impact on trendability and generally degrade RMSE, particularly at higher α values and in the CGGD setting. This behavior can be explained again by the degradation characteristics shown in Fig. 4, where the vibration energy remains low and fairly constant until the final portion of the run, causing the RMS signal to only become informative near end of life. When α is small and the supervised MSE term remains dominant, a modest RMSE improvement is observed for Bearing1-4. In this case, the late-life increase in RMS provides slight corrective guidance, as illustrated in Fig. 3d. However, this benefit diminishes as α increases and the energy penalty begins to dominate the optimization. In the CGGD-energy formulation, where no supervised objective is present, performance deteriorates further. Including MSE reduces the large RMSE but still results in no improvements over the baseline metrics. As shown in Fig. 3e, the model largely follows the RMS pattern itself, which does not serve as a reliable proxy for RUL in this dataset as the decrease in RUL would be noticed far too late. The observed increase in robustness can also be attributed to this behavior: because the RMS signal remains relatively stable for most of the run, the resulting RUL predictions are also stable prior to the final decline, causing the robustness metric to be overestimated. Overall, these results suggest that the energy constraint is poorly supported by the PHM2012 degradation characteristics. For constraints based on signal RMS to be beneficial, the data must exhibit a more consistent degradation trend.

Table 3. Performance metrics per model for the private dataset. All bold values indicate the best performance per metric and are statistically significant compared to the baseline using a paired statistical test ($p < 0.05$).

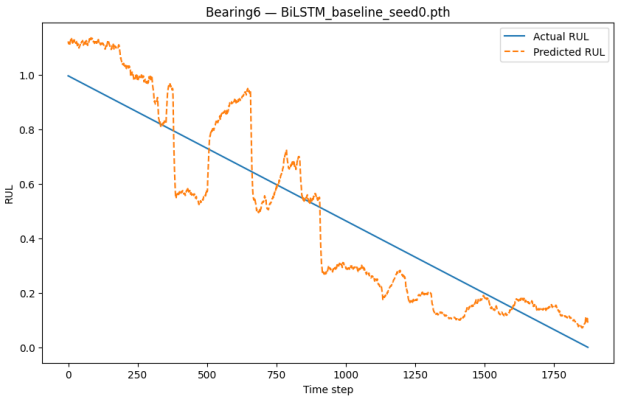
Model α	**Bearing 2020-02-14**			**Bearing 2020-02-18**		
	RMSE	Trendability	Robustness	RMSE	Trendability	Robustness
Baseline	0.156 ± 0.006	-0.784 ± 0.038	0.723 ± 0.013	0.135 ± 0.010	-0.953 ± 0.008	0.743 ± 0.018
Mono 5	0.138 ± 0.004	-0.878 ± 0.006	0.777 ± 0.015	0.107 ± 0.006	$\mathbf{-0.969 \pm 0.005}$	0.825 ± 0.016
Mono 20	0.148 ± 0.006	-0.899 ± 0.011	0.850 ± 0.018	0.114 ± 0.009	$\mathbf{-0.969 \pm 0.004}$	0.897 ± 0.014
Mono 50	0.165 ± 0.010	-0.932 ± 0.015	$\mathbf{0.885 \pm 0.012}$	0.143 ± 0.010	-0.967 ± 0.002	$\mathbf{0.931 \pm 0.008}$
Mono CGGD	0.247 ± 0.037	$\mathbf{-0.951 \pm 0.010}$	0.739 ± 0.043	0.223 ± 0.032	$\mathbf{-0.969 \pm 0.005}$	0.663 ± 0.035
Energy 100	0.147 ± 0.005	-0.835 ± 0.026	0.729 ± 0.006	0.134 ± 0.010	-0.950 ± 0.007	0.738 ± 0.011
Energy 1000	$\mathbf{0.134 \pm 0.004}$	-0.882 ± 0.007	0.772 ± 0.013	$\mathbf{0.097 \pm 0.004}$	-0.951 ± 0.004	0.807 ± 0.016
Energy 10000	0.145 ± 0.010	-0.883 ± 0.005	0.806 ± 0.025	0.100 ± 0.015	-0.958 ± 0.004	0.865 ± 0.021
Energy CGGD	0.250 ± 0.022	-0.870 ± 0.008	0.678 ± 0.037	0.200 ± 0.027	-0.964 ± 0.003	0.759 ± 0.047

5.2. Private dataset

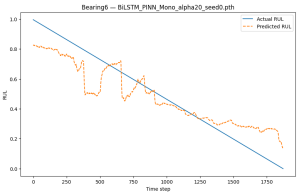
The results for the private dataset are summarized in Table 3, the prediction trajectories for Bearing2020-02-18 are shown in Fig. 5. The dataset exhibits different degradation characteristics compared to PHM2012, mainly evolving in a more gradual manner. This is clearly reflected in the spectrogram in Fig. 6, where broadband vibration energy increases steadily with a pronounced shift around time step 1000 rather than only near failure. This results in noticeably different behavior compared to PHM2012.

The monotonic loss exhibits behavior consistent with its performance on the PHM2012 dataset. At lower constraint weights, RMSE typically improves while trendability and robustness increase. As the weight grows, the same trade-off reappears: physical consistency continues to improve, but at the cost of higher RMSE. For Bearing2020-02-18, this effect appears less pronounced in Table 3, as the baseline predictions are already largely monotonic. Nevertheless, Fig. 5b shows that small, physically inconsistent local arcs remain. Such residual violations can persist because enforcing the constraint during training does not guarantee perfect generalization to unseen test data. Distributional differences between training and test sets may still induce minor departures from monotonicity. Even so, the magnitude of these inconsistencies is consistently reduced across all constraint-based models.

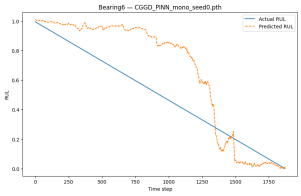
The monotonic CGGD model again achieves the highest trendability, but its behavior differs notably from the PHM2012 results. In PHM2012, where vibration energy remained largely flat until the final stage of degradation, monotonic CGGD produced a smooth, near-linear trajectory closely aligned with the ideal RUL target. In contrast, the private dataset exhibits a clear mid-life rise in vibration energy, as shown in the spectrogram in Fig. 6, which is reflected in the extracted training features. Because CGGD updates are triggered solely by constraint violations and do not explicitly enforce linearity, mid-life feature variations induce additional corrective updates. As a result, the predicted RUL drops more steeply halfway through the run (Fig. 5c). Although the trajectory remains physically plausible and fully eliminates the minor arcs observed in other models, it deviates from the ideal linear trend, resulting in higher RMSE and reduced robustness.



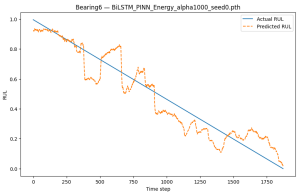
(a) Baseline



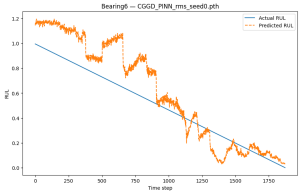
(b) Mono 5.0



(c) Mono CGGD



(d) Energy 5.0



(e) Energy CGGD

Figure 5. Effect of monotonic and energy physical loss functions on predictions for Bearing2020-02-18 from the private dataset. RUL values are normalized between 0 and 1. One time step is equal to 6 seconds.

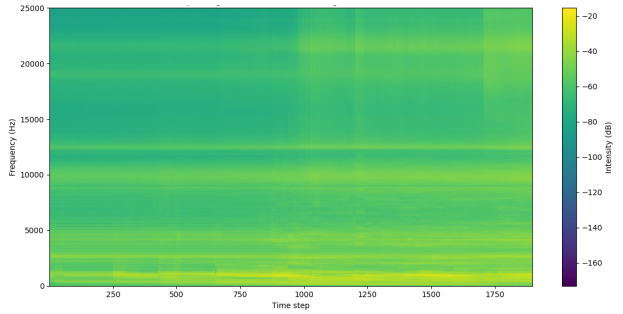


Figure 6. Spectrogram of the raw accelerometer signal for Bearing2020-02-18 in the private dataset. One time step is equal to 6 seconds.

The energy-based loss performs substantially better on the private dataset than on PHM2012. Due to the different degradation mechanics in this setup, the RMS increases in a smoother and more consistently monotonic pattern that closely matches the expected evolution of vibration energy as damage progresses. This provides the constraint with clear directional information and enables it to act as an effective regularizer. As illustrated in Fig. 5d, the resulting predictions follow the target RUL more closely across the entire run compared to the baseline. Consequently, the energy-based loss achieves the best RMSE improvements among all models while simultaneously enhancing trendability and robustness, in contrast to the more limited effect observed on PHM2012.

The energy CGGD model (Fig. 5e) produces predictions that closely resemble the baseline, as the RMS signal increases in a smooth and consistently monotonic manner that already aligns with the target RUL trend. Because of this, the trendability using energy CGGD also increases. However, the other quantitative metrics show mixed effects: RMSE increases and robustness varies across bearings. This behavior can be attributed to the ZeroLoss formulation, which strictly follows the RMS-driven constraint direction and relies solely on boundary conditions, without an explicit mechanism to correct deviations toward the true RUL target. Yet, in practice, adding MSE did not improve any metrics over the baseline.

6. CONCLUSION

This work investigated how constraint integration can enhance the physical reliability of data-driven RUL prediction. By comparing physics-guided loss functions with CGGD, we examined how different constraint mechanisms influence predictive behavior beyond traditional accuracy-based metrics. The study shows that incorporating even relatively simple physical assumptions can meaningfully reshape degradation trajectories.

Overall, constraint-based approaches are capable of enforcing more physically consistent RUL predictions, but the extent of their effectiveness depends on how well the imposed assumptions interact with the underlying degradation dynamics. Monotonicity constraints consistently enhance physical plausibility, whereas energy-based constraints prove more sensitive to vibration progression. CGGD offers strong guarantees of constraint satisfaction without requiring loss-weight tuning, but its predictive accuracy is highly context-dependent, achieving the best performance on one dataset and the weakest on the other. These findings reinforce that physically consistent predictions are a multi-faceted objective: improvements in trendability and robustness may not always coincide with improvements in accuracy, and vice versa. Consequently, meaningful evaluation of RUL models must consider physical consistency metrics alongside MSE to ensure a fair assessment of model behavior.

The effects of the various constraint models on the main objectives can be summarized as follows:

- **Trendability.** Both physics-guided losses and CGGD improve monotonic RUL behavior relative to baseline models. Among them, CGGD is the most consistent, enforcing strictly decreasing trajectories and generally achieving the strongest increases in trendability.
- **Accuracy (RMSE).** RMSE performance proved highly context-dependent, which may reflect the fact that the ground-truth trajectory is itself a simplified approximation of degradation when compared with the richer patterns captured in the spectrograms. Across all experiments, CGGD and/or physics-guided losses still achieved accuracy comparable to or better than the baseline.
- **Robustness.** Physics-guided losses generally act as effective regularizers. By balancing accuracy and physics, they reduce sensitivity to noise and limit overfitting to short-term fluctuations. As a result, they tend to produce consistently smoother and more stable degradation trajectories than both the baseline and CGGD.

These observations demonstrate that no single method universally optimizes all objectives and that the optimal constraint must be matched carefully to the system at hand.

Future work offers several promising possibilities for extending this study. One major direction is to further attempt CGGD integration with supervised objectives such as RMSE, with results suggesting potential improvement, though initial tests have not yet achieved this. Another direction is to explore additional constraints and how they interact when applied simultaneously as well as which ones are activated most. Extensions could also consider alternative formulations, such as enforcing constraints over all pairs of time steps or over large time intervals. Further work may also look at relaxing the current assumptions to better accommodate more complex operating conditions (e.g. self-healing, changing operating conditions). Finally, specifically for physics-guided loss functions, adaptive weighting strategies could be applied to optimize the loss

weight automatically. Since the proposed constraints express general physical assumptions rather than PINN-style physics modeling, they can be easily transferred to other RUL prognostics tasks such as battery or gearbox degradation prediction. The same evaluation metrics can also be directly applied.

ACKNOWLEDGMENT

This research was supported by Flanders Make, the strategic research centre for the manufacturing industry, through the SBO project SAFOS and by the Flemish Government under the Flanders AI Research Program.

REFERENCES

Biggio, L., & Kastanis, I. (2020). Prognostics and Health Management of Industrial Assets: Current Progress and Road Ahead. *Frontiers in Artificial Intelligence*, *3*, 578613. doi: 10.3389/frai.2020.578613

Carino, J. A., Zurita, D., Delgado, M., Ortega, J. A., & Romero-Troncoso, R. J. (2015). Remaining useful life estimation of ball bearings by means of monotonic score calibration. In *2015 ieee international conference on industrial technology (icit)* (p. 1752-1758). doi: 10.1109/ICIT.2015.7125351

Chen, J., & Liu, Y. (2021). Probabilistic physics-guided machine learning for fatigue data analysis. *Expert Systems with Applications*, *168*, 114316. doi: https://doi.org/10.1016/j.eswa.2020.114316

Chen, Y., Rao, M., Feng, K., & Zuo, M. J. (2022). Physics-Informed LSTM Hyperparameters Selection for Gearbox Fault Detection. *Mechanical Systems and Signal Processing*, *171*, 108907. doi: 10.1016/j.ymssp.2022.108907

Chennana, A., Megherbi, A. C., Bessous, N., Sbaa, S., Teta, A., Belabbaci, E. O., ... Agajie, T. F. (2025). Vibration signal analysis for rolling bearings faults diagnosis based on deep-shallow features fusion. *Scientific Reports*, *15*, 9270. doi: 10.1038/s41598-025-93133-y

Cleveland, W. S., & Devlin, S. J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, *83*(403), 596–610.

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5), 602-610. (IJCNN 2005) doi: https://doi.org/10.1016/j.neunet.2005.06.042

He, Y., Su, H., Zio, E., Fan, L., & Zhang, J. (2022). A Physics-Informed Training Approach for Data-Driven Method in Remaining Useful Life Estimation. In *2022 6th International Conference on System Reliability and Safety (ICSRS)* (p. 500-504). doi: 10.1109/ICSRS56243.2022.10067722

Herwig, N., Borghesani, P., Smith, W., & Peng, Z. (2025). Signal processing- and physics-informed neural network for explainable bearing condition monitoring. *Mechanical Systems and Signal Processing*, *235*, 112925. doi: https://doi.org/10.1016/j.ymssp.2025.112925

Li, H., Zhang, Z., Li, T., & Si, X. (2024). A review on physics-informed data-driven remaining useful life prediction: Challenges and opportunities. *Mechanical Systems and Signal Processing*, *209*, 111120. doi: https://doi.org/10.1016/j.ymssp.2024.111120

Liao, X., Chen, S., Wen, P., & Zhao, S. (2023). Remaining useful life with self-attention assisted physics-informed neural network. *Advanced Engineering Informatics*, *58*, 102195. doi: https://doi.org/10.1016/j.aei.2023.102195

Lu, H., Pavan Nemani, V., Barzegar, V., Allen, C., Hu, C., Laflamme, S., ... Zimmerman, A. T. (2023). A physics-informed feature weighting method for bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *191*, 110171. doi: https://doi.org/10.1016/j.ymssp.2023.110171

Lu, W., Wang, Y., Zhang, M., & Gu, J. (2024). Physics guided neural network: Remaining useful life prediction of rolling bearings using long short-term memory network through dynamic weighting of degradation process. *Engineering Applications of Artificial Intelligence*, *127*, 107350. doi: https://doi.org/10.1016/j.engappai.2023.107350

Luo, J., & Zhang, X. (2022, January). Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction. *Applied Intelligence*, *52*(1), 1076–1091. doi: 10.1007/s10489-021-02503-2

Meire, M., Brijder, R., Dekkers, G., & Karsmakers, P. (2022, 10). Accelerometer-Based Bearing Condition Indicator Estimation Using Semi-Supervised Adaptive DSVDD. *Annual Conference of the PHM Society*, *14*. doi: 10.36001/phmconf.2022.v14i1.3173

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., & Varnier, C. (2012, 06). PRONOSTIA: An experimental platform for bearings accelerated degradation tests. In (p. 1-8).

Park, H. J., Kim, N. H., & Choi, J.-H. (2024). A robust health prediction using Bayesian approach guided by physical constraints. *Reliability Engineering System Safety*, *244*, 109954. doi: https://doi.org/10.1016/j.ress.2024.109954

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686-707. doi: https://doi.org/10.1016/j.jcp.2018.10.045

Rombouts, W., Van Baelen, Q., & Karsmakers, P. (2025). *Constraint-Guided PINNs: A Constrained Optimization*

Approach (Vol. 4125). R. Piskac c/o Redaktion Sun SITE Informatik V RWTH Aachen.

Song, L., Lin, T., Jin, Y., Zhao, S., Li, Y., & Wang, H. (2024, jun). Advancements in bearing remaining useful life prediction methods: a comprehensive review. *Measurement Science and Technology*, *35*(9), 092003. doi: 10.1088/1361-6501/ad5223

Tefera, Y., Van Baelen, Q., Meire, M., Luca, S., & Karsmakers, P. (2025). *Constraint-Guided Learning of Data-Driven Health Indicator Models: An Application on Bearings* (Vol. 16) (No. 2). The Prognostics and Health Management Society.

Van Baelen, Q., & Karsmakers, P. (2023, November). Constraint guided gradient descent: Training with inequality constraints with applications in regression and semantic segmentation. *Neurocomputing*, *556*, 126636. doi: 10.1016/j.neucom.2023.126636

Wang, S., Yan, L., Du, S., Li, S., & Chen, X. (2025, apr). Bearing prognostics and health management based on hybrid physical mechanism and data models: a systematic review. *Measurement Science and Technology*, *36*(5), 052002. doi: 10.1088/1361-6501/adcce4

Wu, F., Wu, Q., Tan, Y., & Xu, X. (2024). Remaining Useful Life Prediction Based on Deep Learning: A Survey. *Sensors*, *24*(11). doi: 10.3390/s24113454

Youness, G., & Aalah, A. (2023). An Explainable Artificial Intelligence Approach for Remaining Useful Life Prediction. *Aerospace*, *10*(5), 474. doi: 10.3390/aerospace10050474

Yucesan, Y. A., & Viana, F. A. (2022). A hybrid physics-informed neural network for main bearing fatigue prognosis under grease quality variation. *Mechanical Systems and Signal Processing*, *171*, 108875. doi: https://doi.org/10.1016/j.ymssp.2022.108875

Zhang, B., Zhang, L., & Xu, J. (2016). Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings. *Quality and Reliability Engineering International*, *32*(2), 547-554. doi: https://doi.org/10.1002/qre.1771

Zhu, Y., Cheng, J., Liu, Z., Zou, X., Wang, Z., Cheng, Q., ... Tao, F. (2024). Remaining Useful Life Prediction Approach Based on Data Model Fusion: An Application in Rolling Bearings. *IEEE Sensors Journal*, *24*(24), 42230-42244. doi: 10.1109/JSEN.2024.3477489