IISE ANNUAL
CONFERENCE & EXPO 2026

A conditional Diffusion Model for Robust Time-to-Failure Prediction under Complex Degradation Processes

Presenter: Donghyun Ko*

* Edward P. Fitts Department of Industrial and Systems Engineering, NC State University, Raleigh, NC, USA

Keywords: Remaining Useful Life (RUL); Time-to-Failure (TTF); Degradation signals; Generative time-series modeling; Diffusion models; prognostics and health management (PHM)

IISE ANNUAL CONFERENCE & EXPO 2026

A Conditional Diffusion-Based Generative AI Model for Industrial Predictive Analytics

Presenter: Donghyun Ko*

* Edward P. Fitts Department of Industrial and Systems Engineering, NC State University, Raleigh, NC, USA

Keywords: Remaining Useful Life (RUL); Time-to-Failure (TTF); Degradation signals; Generative time-series modeling; Diffusion models; prognostics and health management (PHM)

Backgrounds

Forecasting RUL / TTF is a key enabler of predictive maintenance and system reliability in manufacturing

⚠️ **No dedicated degradation sensor exists** — signals must be engineered from raw sensor data reflecting aging behavior.

① Sensor Selection

Select only failure-informative sensors

- Group LASSO, Elastic Net, and etc.

② Feature Extraction

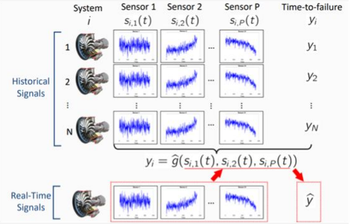
Reduce dimension of the selected sensors via PCA, or FPCA

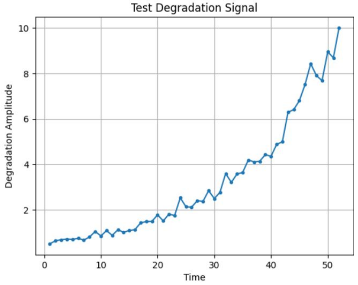
- Top-K principal scores are used as extracted features

③ Signal Fusion

Fuse the selected features

- Degradation data typically shows 1D exponential increasing pattern with noise
- Complex system shows non-monotonic, or multimodal degradation pattern (e.g., thermal cycling, load/speed change…)





Traditional Approaches

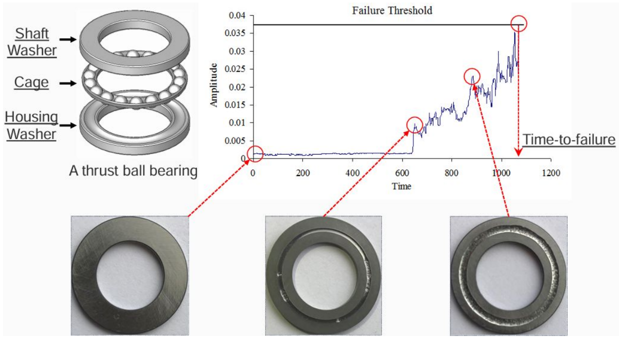
Model-Based

- Explicitly model stochastic degradation; Bayesian updating of parameters (Gebraeel et al. 2005)

Data-Driven

- PCA for dimension reduction, followed by regression (e.g., log-normal regression)

Benchmark 1: Bayesian Updating



Objective: To model the functional form of the degradation process

- **Candidates: 'Exponential + i.i.d' vs. 'Exponential + Brownian Motion(BM)'**
  - Degradation is assumed to follow an exponential form: $S(t_i) = \phi + \theta \exp\left(\beta t_i + \varepsilon(t_i) - \frac{\sigma^2}{2}\right)$

    Where:
    - ϕ: A constant (typically the baseline level)
    - θ: A log-normal random variable representing the initial state of the component
    - β: Degradation rate (normally distributed)
  - **i.i.d Model:** This model uses a log-transformation to express the degradation process in linear form. Then, linear regression can be applied to estimate the initial condition and degradation rate of each component

    $\varepsilon(t_i)$: Observation error, assumed to be i.i.d. normal $N(0, \sigma^2)$ ➡ $L(t_i) = \ln(S(t_i) - \phi) = \ln \theta + \beta t_i + \varepsilon(t_i) - \frac{\sigma^2}{2}$
  - **BM Model:** Same modeling, yet with assumption that errors are from Brownian motion error.

    It is suitable for modeling degradation processes with time-dependent noise.

    $\varepsilon(t) = \sigma W(t)$ where $W(t)$ is a standard Brownian motion ➡ $L(t) = \theta' + \beta' t + \varepsilon(t)$

    Parameters of the distribution

    Observed degradation signal amplitude at time step 't'

Benchmark 1: Bayesian Updating (Cont')

➢ 3 Steps: Assume prior distribution - Posterior update - Derive RUL distribution

- Step 1. **Assume the prior distribution** of the parameters in the two candidates as 'Gaussian distribution'

  (Can be estimated by training data) $\theta' \sim N(\mu_{\theta'}, \sigma_{\theta'}^2), \quad \beta \sim N(\mu_\beta, \sigma_\beta^2)$ ➡ $\pi(\theta') = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{(\theta' - \mu_{\theta'})^2}{2\sigma_0^2}\right\}$ $\pi(\beta) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(\beta - \mu_\beta)^2}{2\sigma_1^2}\right\}$

  "Prior"

- Step 2. **Estimate posterior distribution** of the parameters by combining priors and observed (test)data
  - i.i.d model: $L(t_i) = \theta' + \beta t_i + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$ ➡ $L(t_i) \mid \theta', \beta \sim N\left(\theta' + \beta t_i,\, \sigma^2\right)$ ➡ $f(L(t_1), \ldots, L(t_k) \mid \theta', \beta) = \prod_{i=1}^{k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{[L(t_i) - \theta' - \beta t_i]^2}{2\sigma^2}\right\} \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\left[L(t_i) - \theta' - \beta t_i\right]^2\right\}$

    "Likelihood"

    RUL regression model

    By bayes' theorem, $p(\theta', \beta \mid \{L(t_i)\}) \propto f(\{L(t_i)\} \mid \theta', \beta)\, \pi(\theta')\, \pi(\beta)$

    ➡ $p(\theta', \beta \mid \{L(t_i)\}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\left[L(t_i) - \theta' - \beta t_i\right]^2 - \frac{(\theta' - \mu_{\theta'})^2}{2\sigma_0^2} - \frac{(\beta - \mu_\beta)^2}{2\sigma_1^2}\right\}$

    "Posterior"

    ➡ Posterior is a bivariate normal distribution: $(\theta', \beta) \mid \{L(t_i)\} \sim \mathcal{N}\left((\mu_{\theta'}^*, \mu_\beta^*), \Sigma^*\right)$

    "This posterior is updated each time new sensor data is received!"

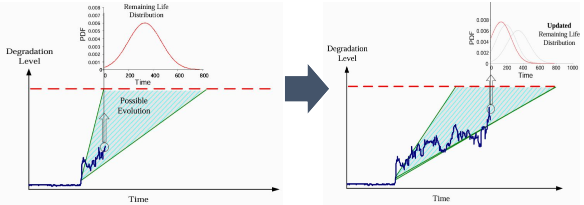
- Step 3. **Derive the Remaining Useful Life(RUL) Distribution**
  - RUL 'T' is the time after the current observation t_k such that $L(t_k + T) = D$

    $L(t_k + T) = \theta' + \beta(t_k + T) + \varepsilon(t_k + T)$ where $\varepsilon(t_k + T) \sim \mathcal{N}(0, \sigma_\varepsilon^2(T))$
    - In the **i.i.d. error model**: $\sigma_\varepsilon^2(T) = \sigma^2$ (constant).
    - In the **Brownian motion model**: $\sigma_\varepsilon^2(T) = \sigma^2(t_k + T)$

    Thus, conditionally:

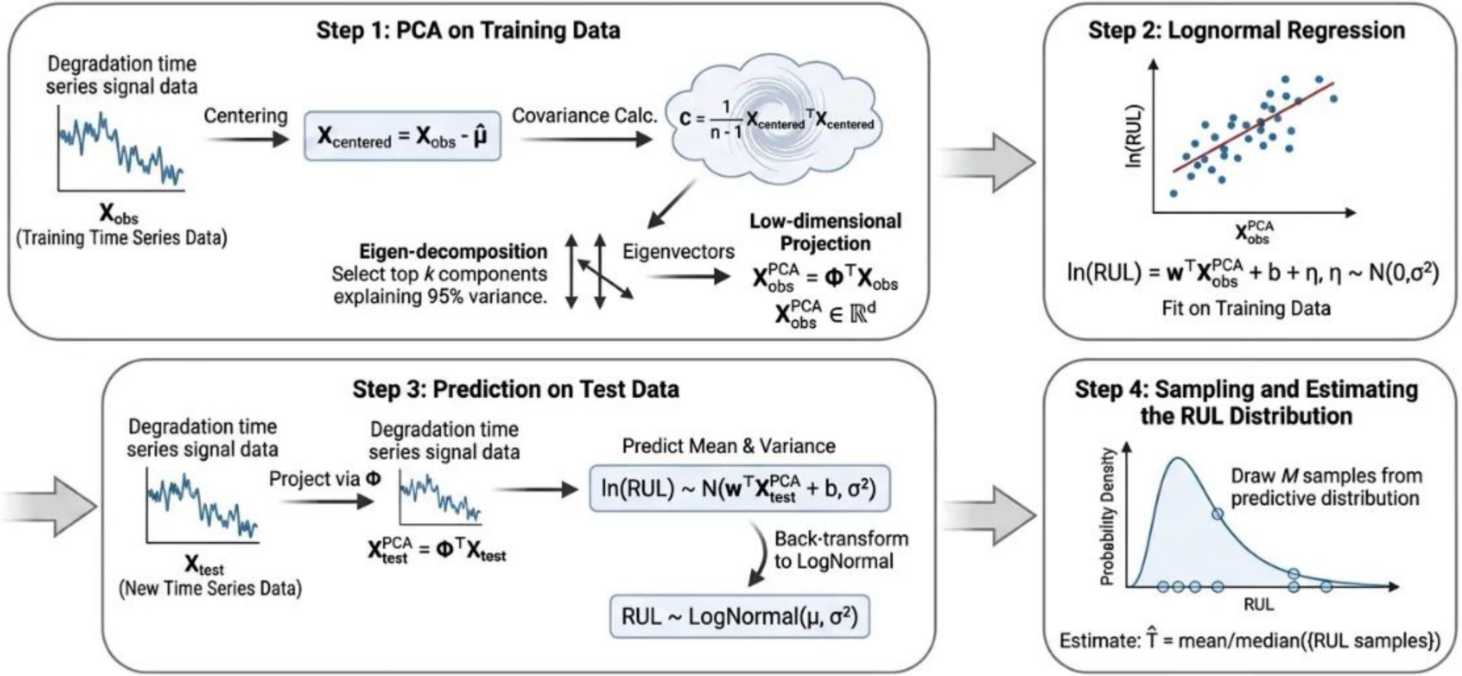
    $L(t_k + T) \sim \mathcal{N}(\mu_{\theta'} + \mu_\beta(t_k + T), \sigma_{\text{post}}^2(t_k + T))$
    - For the **i.i.d. case**: $\sigma_{\text{post}}^2(t_k + T) = \sigma_{\theta'}^2 + (t_k + T)^2\sigma_\beta^2 + \sigma^2$
    - For the **Brownian motion case**: $\sigma_{\text{post}}^2(t_k + T) = \sigma_{\theta'}^2 + (t_k + T)^2\sigma_\beta^2 + \sigma^2(t_k + T)$

    Define: $\mu(t_k + T) = \mu_{\theta'} + \mu_\beta(t_k + T)$ ➡ Then: $L(t_k + T) \sim \mathcal{N}(\mu(t_k + T), \sigma^2(t_k + T))$



  - Using the posterior distribution, $L(t_k + T)$ is treated as a normally distributed variable, leading to a CDF such that: $P\{T \le t \mid L(t_1), \ldots, L(t_k)\} = P\{L(t_k + t) \ge D \mid L(t_1), \ldots, L(t_k)\}$
  - PDF of RUL: $f_T(t \mid \{L(t_i)\}) = \frac{d}{dt}F_T(t) \approx \phi(g(t)) \cdot g'(t)$ Where:
    - $\phi(\cdot)$ is the standard normal PDF,
    - $g'(t)$ is the derivative of $g(t)$ with respect to t

Benchmark 2: PCA + Lognormal Regression



Research Gap & Our Approach

Gaps in Existing Literature

No systematic comparison with classical baselines
e.g., Bayesian updating, PCA + Lognormal

↓

Many studies use curated datasets with smooth, near-monotone degradation
Real-world degradation is often noisier and irregular

↓

Uncertainty about robustness under complex, abrupt signal behaviors
Generalization to unseen conditions remains unclear

Gaps addressed by this work →

This Work Addresses Two Critical Questions

1 Diffusion Model | Bayesian Updating

When and by how much does Diffusion outperform Bayesian Updating (IID / BM)?

Bayesian updating limitation

$y = f(x; \theta)$

- Assumes a fixed functional form
- Does not truly learn from data
- If real signals deviate, predictions become unreliable

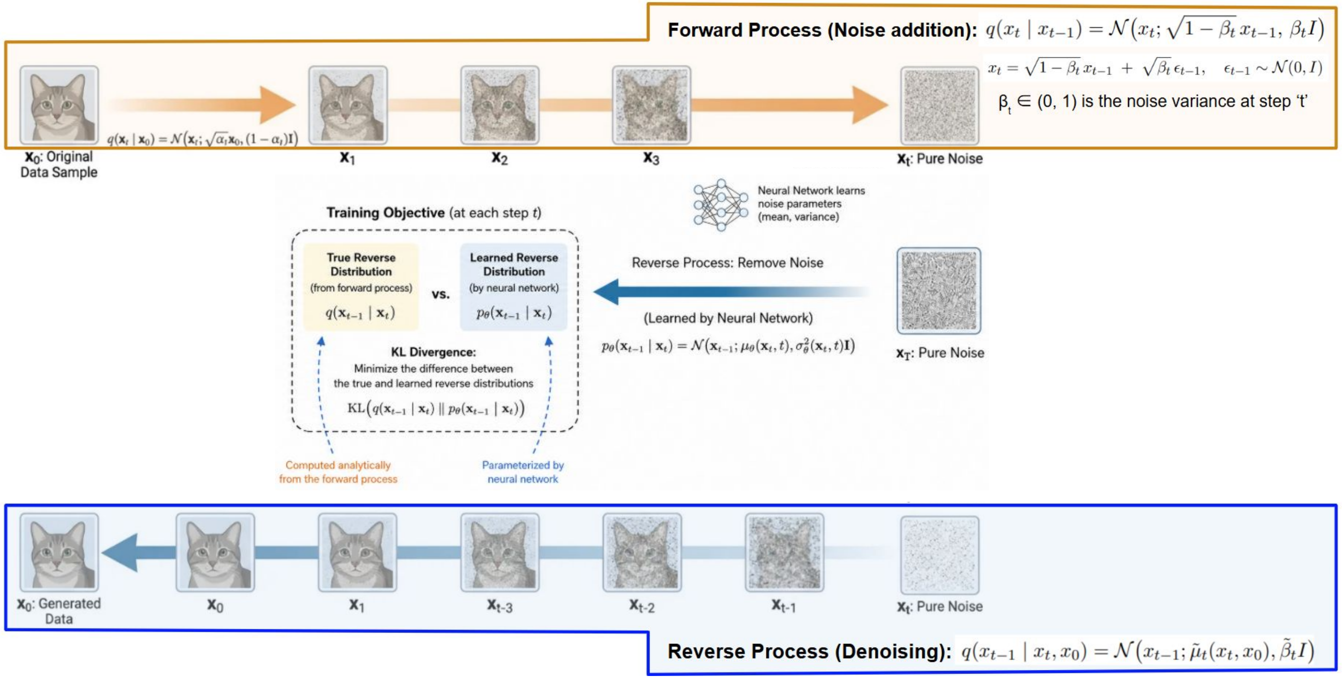
2 Diffusion Model | PCA + Lognormal

When and by how much does Diffusion outperform PCA + Lognormal regression?

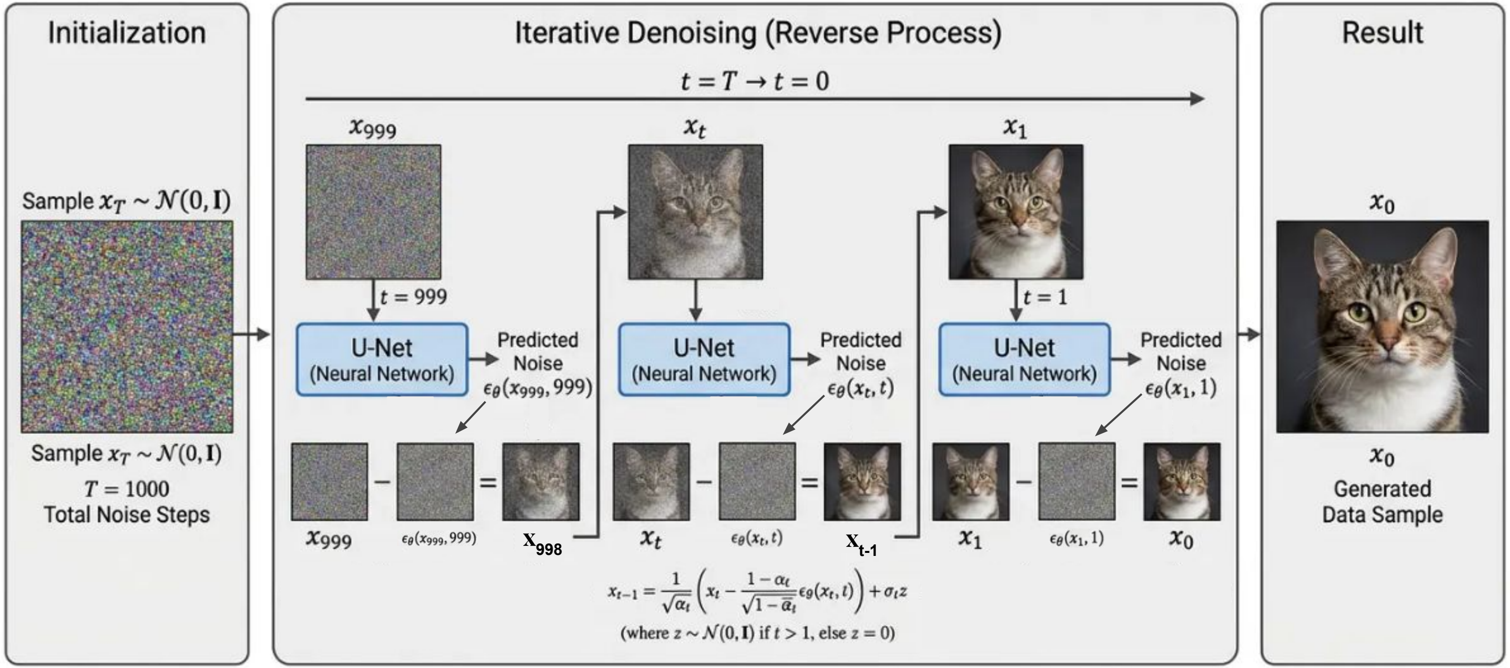
PCA + Lognormal limitation

- PCA projects into a linear subspace
- Complex degradation may lie on a nonlinear manifold
- Projection can lose critical information
- Poor prediction may result

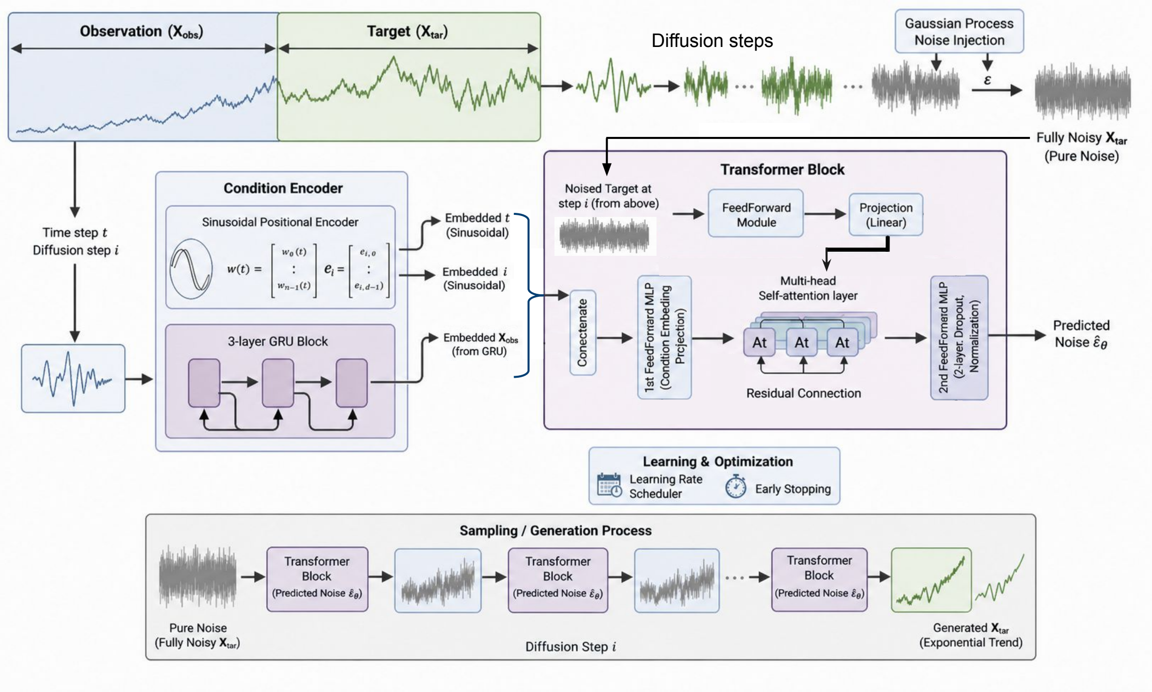
Diffusion Model Fundamentals: Forward & Reverse Process



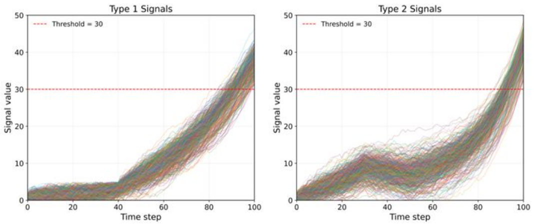
Diffusion Model Fundamentals: Sampling Process



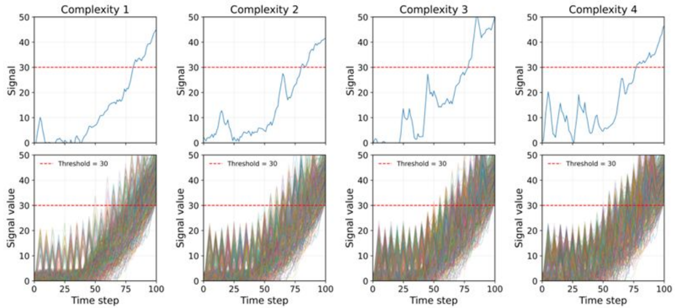
Conditional Diffusion Model & Architecture



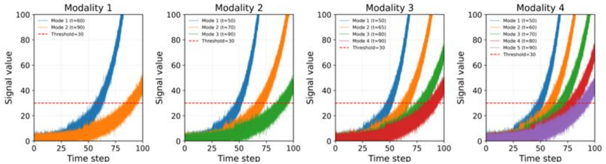
Dataset



- ❖ **Level 1** — Typical degradation patterns
  - ➢ **Type 1 (Monotone)**: Exponential growth perturbed by Brownian-motion noise
  - ➢ **Type 2 (Non-Monotone)**: Initial exponential growth → short pullback → exponential rise again



- ❖ **Level 2** — Complex and irregular degradation patterns
  - ➢ **Complexity 1–4**: Randomly inject 'K' localized bumps on Type 1
    - ■ K = 1, 2, 3, 4 for Complexity 1–4; bump amplitudes ~ U[10, 20]
    - ■ Mimics thermal spikes, shocks, or control instabilities
  - ➢ **Modality 1–4:** Trajectories generated under 'M' operating regimes
    - ■ M = 1, 2, 3, 4 for Modality 1–4 (number of extra regimes)
    - ■ Produces multimodal TTF distributions with heteroscedastic behavior



- ❖ 10-Fold Cross-Validation; sample sizes N = 100, 1,000, 5,000, 10,000
  - ➢ 4 observation ratios tested: 2:8, 4:6, 6:4, 8:2 (Observed : Remaining life)
  - ➢ Metrics: RMSE and MAPE of 'predicted mean RUL' vs. 'Ground Truth'

Design of Experiments

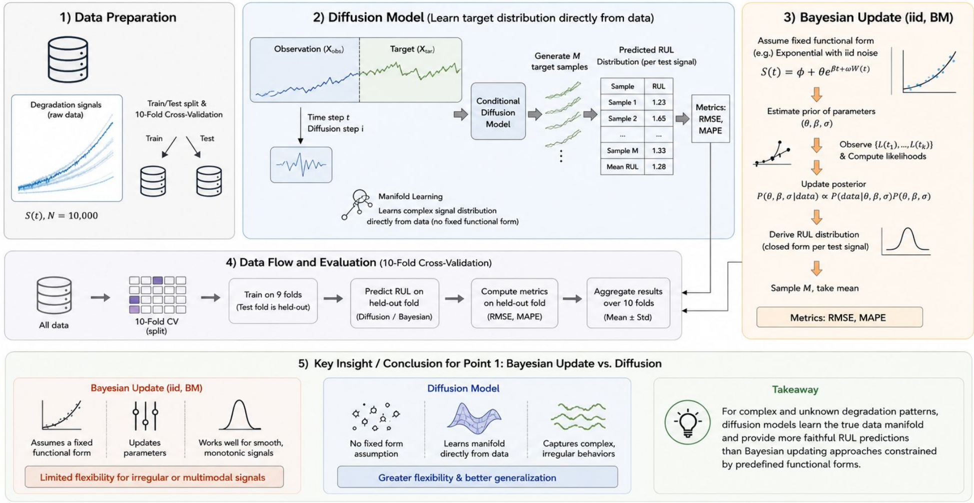
❖ **Point 1** — Bayesian Updating vs. Diffusion

Bayesian (IID / BM):

- ❑ Estimate Gaussian priors from training fold
- ❑ Observe $\{L(t_1),\ldots,L(t_k)\}$ per test signal; update posterior analytically
- ❑ Sample ‘M’ RUL values per test signal via inverse-CDF; take mean

Diffusion:

- ❑ Train Conditional Diffusion Model on training fold per ratio
- ❑ Use observed signal as condition; generate ‘M’ target signals; compute mean RUL



Design of Experiments (Cont’)

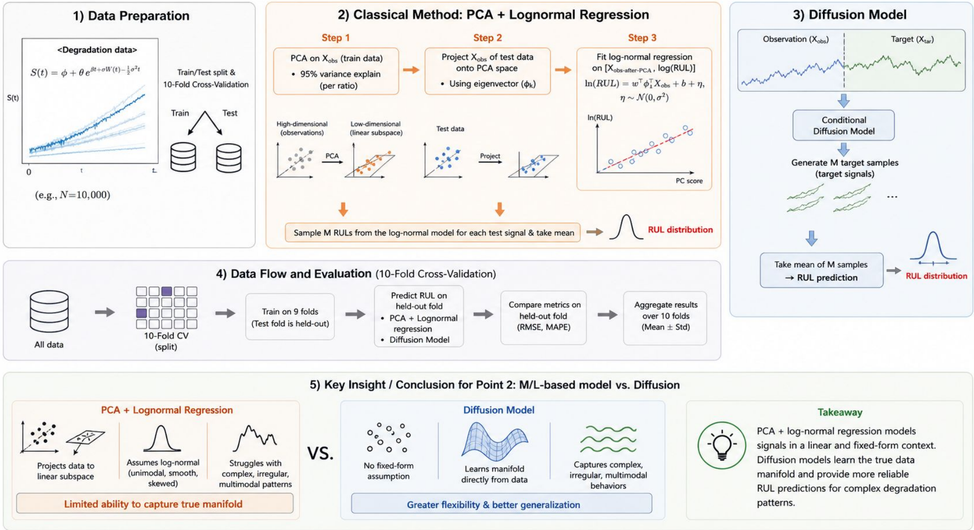
❖ **Point 2** — PCA + Lognormal vs. Diffusion

PCA + Lognormal:

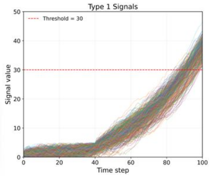
- ❏ PCA on training ‘X_obs’ with 95% variance; fit log-normal regression model
- ❏ Project test ‘X_obs’ onto PCA space; sample ‘M’ RUL values; take mean

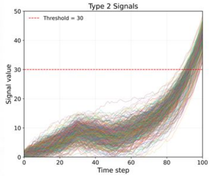
Diffusion:

- ❏ Train Conditional Diffusion Model on training fold per ratio
- ❏ Use observed signal as condition; generate ‘M’ target signals; compute mean RUL



Results: Diffusion vs. Bayesian Updating





Focus of Analysis

Model performance comparison under varying signal characteristics

- ➤ To find where **Diffusion-based RUL/TTF** model perform better than SOTA models and how much better?

Assessing how performance evolves as more of the signal is revealed

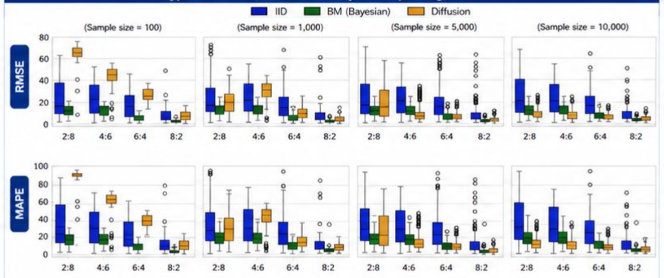
Type 1: Diffusion outperformed Bayesian updating as data size increases (sample size ≥ 5,000)

- **Bayesian update:** competitive performance under limited training conditions (sample size < 5,000)
  - Due to the alignment between their parametric assumption and the structure of the data

Type 2: Diffusion demonstrates clear / robust superiority across all observation ratios (including 8:2)

- Even with smaller data (≥ 1,000), Diffusion consistently yields lower RMSE and MAPE than BM model
  - This performance gap widens as the training set becomes larger
  - Highlighting the ability of diffusion to model complex degradation dynamics
- BM model plateaus due to its heavy reliance on predefined priors

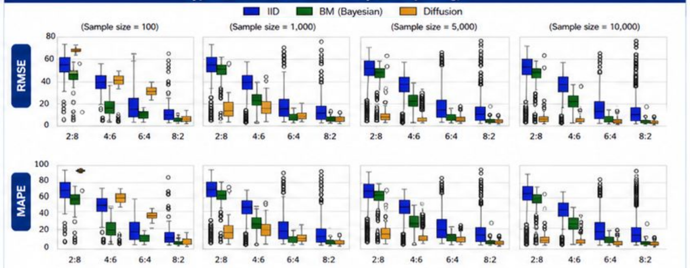
Type 1 Results (Diffusion vs. Bayesian Updating)



Observation Ratio (Observation : Target)

As sample size increases (100 → 10,000) → Diffusion shows consistent improvement in both RMSE and MAPE across all observation ratios

Type 2 Results (Diffusion vs. Bayesian Updating)



- Diffusion is consistently better than BM across all ratios (2:8 ~ 8:2)
- Even with smaller data (≥ 1,000), Diffusion shows lower RMSE & MAPE
- The performance gap widens as the training set becomes larger
- BM model plateaus due to heavy reliance on predefined priors

$$\Delta = \frac{BM - Diff}{BM} \times 100\%$$

Setting	ΔRMSE (%) (Mean ± Std)	ΔMAPE (%) (Mean ± Std)
Type 1	10.04 ± 56.89	3.97 ± 68.77
Type 2	63.35 ± 27.08	63.12 ± 27.40

- Aggregate improvements of Diffusion over BM when N ≥ 10,000 (Mean ± Std)

Overall Takeaway

✓ **Type 1:** Diffusion surpasses Bayesian updating as data size increases (especially when sample size ≥ 5,000).

✓ **Type 2:** Diffusion demonstrates clear and robust superiority across all observation ratios, even with limited data, and the advantage grows with more data.

Results: Diffusion vs. PCA + Lognormal

Focus of Analysis

- Compare **Diffusion-based RUL/TTF** with PCA+Lognormal under varying signal characteristics.
- Identify where PCA+Lognormal stays competitive and where **Diffusion** becomes clearly superior.

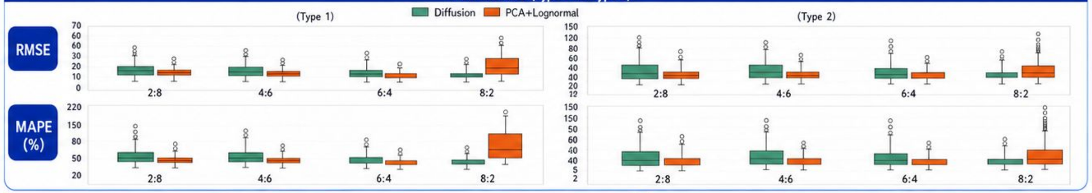
Core Pattern 1: PCA + Lognormal is competitive in simpler settings

- In Type 1–2, performance is strong from 2:8–6:4.
- Threshold crossings are mainly drift-dominated and approximately homoscedastic.

Core Pattern 2: Diffusion wins in harder / noisier settings

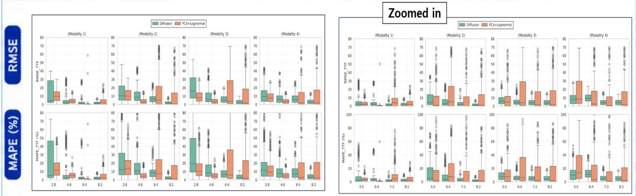
- At 8:2 and under higher modality/complexity, Diffusion achieves lower RMSE and MAPE.
- Near-threshold behavior becomes noise-dominated and heteroscedastic, favoring Diffusion.

Main Results (Type 1 vs. Type 2)



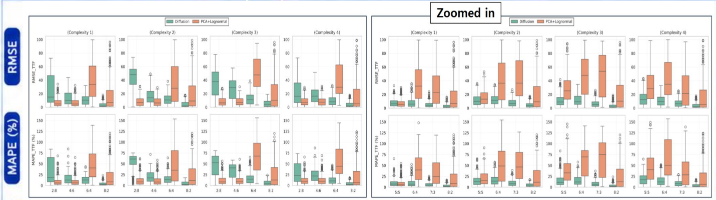
Insight: Diffusion-based models outperform PCA+Lognormal when RUL distributions become noise-dominated.

Modality 1–4 Results



- **Modality 1:** Multi-modality ends near the 6:4 ratio; Noise-dominated: 4:6 and 8:2.
- **Modality 2–4:** Regime switching and sensor drift raise structural variance and introduce between-regime mixtures; Noise-dominated: 5:5 to 8:2.

Complexity 1–4 Results



- **Low–Mid ratios:** Threshold crossings are largely drift-dominated with approximately homoscedastic variability; PCA features with a lognormal TTF fit remain competitive.
- **From 6:4 ratio:** Crossings become noise-dominated and heteroscedastic; Diffusion achieves lower RMSE and MAPE.

Overall Takeaway

✓ Diffusion is most useful when the signal becomes noise-dominated, heteroscedastic, or structurally complex.

✓ PCA+Lognormal remains reasonable in simpler drift-dominated settings, but Diffusion becomes clearly superior as difficulty increases.

Results: Diffusion vs. PCA + Lognormal (Cont')

1 Main Finding

- In both Type 1 and Type 2, observation ratios from 2:8 to 6:4 are largely **drift-dominated** with **near-homoscedastic** variability.
- The **PCA+lognormal baseline** remains competitive in that range.
- Near 8:2, the crossings become **noise-dominated** and **heteroscedastic**.
- **Diffusion attains lower RMSE and MAPE** by modeling the near-threshold dynamics that linear projections with a parametric lognormal fit cannot capture.

2 Type 1 vs. Type 2 at 8:2

At 8:2, Diffusion clearly outperforms PCA+lognormal

About 90% lower error in Type 1 and roughly 32% lower in Type 2.

Setting	ΔRMSE (%)	ΔMAPE (%)
Type 1	90.55 ± 2.10	90.55 ± 2.13
Type 2	33.14 ± 18.51	31.74 ± 19.42

Much larger gain ↑

Smaller gain

3 Across Complexity 1–4

Average improvement over PCA+Lognormal

- Across Complexity 1–4, the average improvement over 6:4–8:2 is **74.47 ± 6.17%** (RMSE) and **75.94 ± 6.50%** (MAPE).
- In the zoomed-in range 5:5–8:2, this rises to **78.62 ± 5.09%** and **79.14 ± 5.75%**.

Improvements on Complexity 1–4

Setting	ΔRMSE (%)	ΔMAPE (%)
Average at 6:4 and 8:2	74.47 ± 6.17	75.94 ± 6.50
Zoom-in (5:5–8:2)	78.62 ± 5.09	79.14 ± 5.75

4 Across Modality 1–4

Average improvement over PCA+Lognormal

- Across Modality 1–4, the corresponding gains are **57.21 ± 19.17%** (RMSE) and **56.53 ± 20.51%** (MAPE) for 6:4–8:2.
- In the zoomed-in range 5:5–8:2, they rise to **59.50 ± 15.71%** and **59.81 ± 17.54%**.

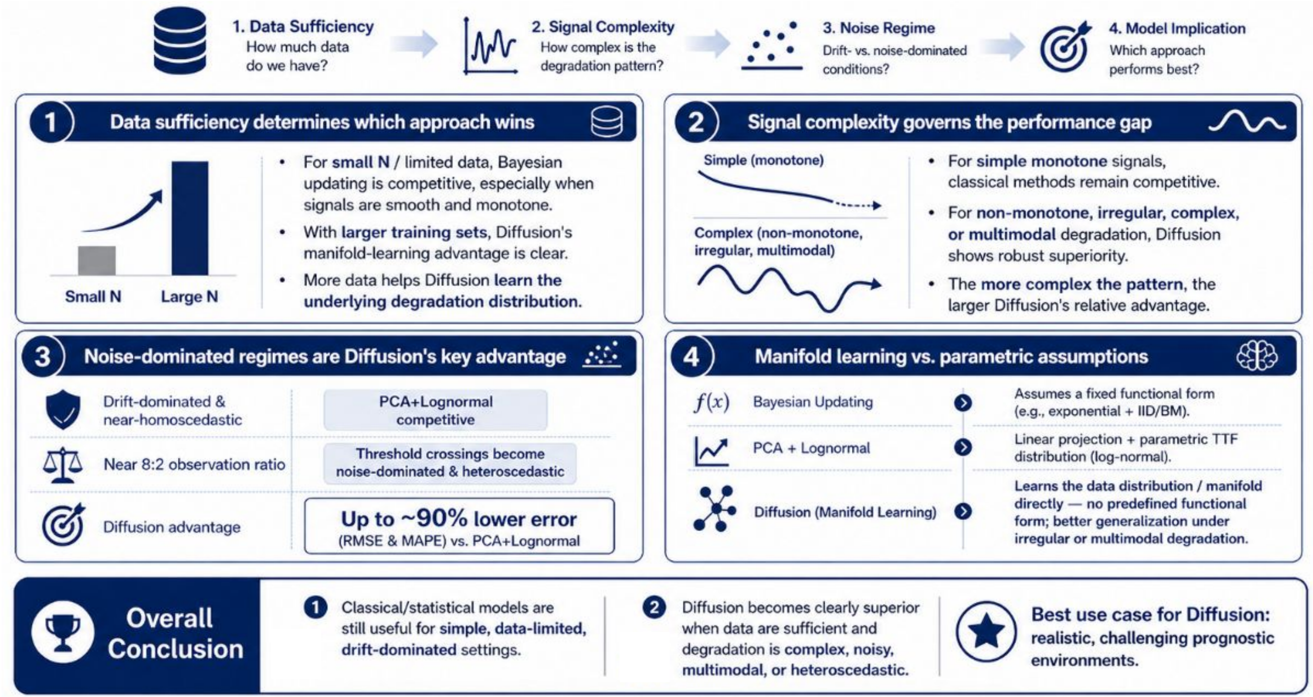
Improvements on Modality 1–4

Setting	ΔRMSE (%)	ΔMAPE (%)
Average at 6:4 and 8:2	57.21 ± 19.17	56.53 ± 20.51
Zoom-in (5:5–8:2)	59.50 ± 15.71	59.81 ± 17.54

Key Takeaways

- Diffusion becomes clearly superior when the regime approaches the noisy near-threshold region.
- The advantage is especially strong in Type 1 and consistently positive across both complexity and modality settings.
- PCA+lognormal remains competitive in simpler, drift-dominated regimes.

Summary & Conclusions



Limitations & Future Research

- Current limitations
  - Evaluation relies entirely on simulated (synthetic) datasets
    - Simulations control complexity but cannot fully replicate real sensor noise
  - Diffusion-based inference is more computationally demanding than classical baselines

- Future research directions
  - Real-world dataset validation
    - IMS bearing data, NASA C-MAPSS, and other benchmark datasets
    - Evaluate generalization under realistic operating conditions
  - Accelerated inference
    - DDIM, progressive distillation, optimized noise schedules
  - Online and federated deployment
    - Streaming inference as new observations arrive
    - Privacy-preserving deployment via federated / meta-learning approaches

IISE ANNUAL CONFERENCE & EXPO 2026

Q & A

Any questions, or comments?

Remember to complete your evaluation for this session within the app!

iise.org/annual #IISEANNUAL2026