Investigating the Reliability of Mechanical Components

In this report, we investigate the reliability and degradation of mechanical components “K”, focusing on accurately modelling the integrity decay over time, predicting future integrity levels (especially with limited data), and identifying key component characteristics affecting degradation. We apply hierarchical Bayesian modelling to achieve these objectives, contrasting its performance with a simple black-box method. We also address additional queries from the client's engineering team regarding modelling choices.

1 Predictive modelling

1.1 Baseline Model

We treat the degradation equations not just as a fixed formula, but as something that has uncertainty - because real-world systems always have variation.

This is the core of the baseline predictive model.

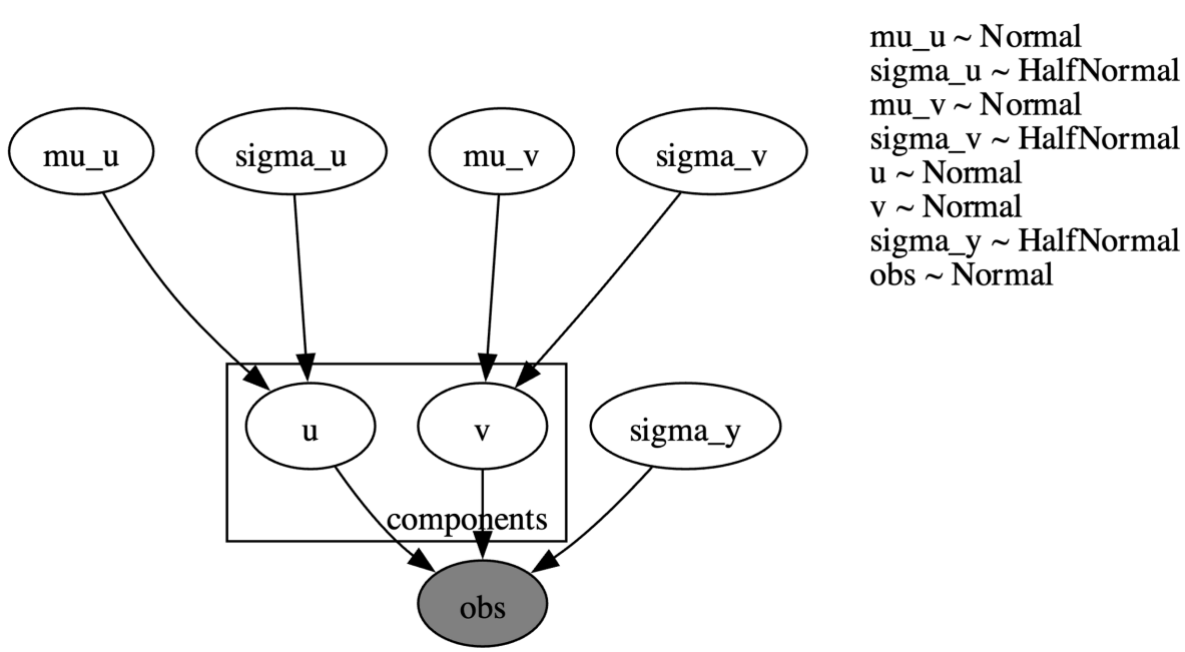
- The observation model is Gaussian, with observation noise sigma_y. The observations are the integrity measurements at each time point.
    - The noise sigma_y is shared across all components and has a half-Gaussian (HalfNormal) hyper-prior.
- The predictor for each component is an exponential decay:

$$\mu(t) = u_i \exp\left(-\frac{v_i t}{100}\right)$$

where t is time in days, u_i is the initial integrity, and v_i is the degradation rate.

- Priors over all u_i and v_i are Gaussian:
    - $u_i \sim N(\mu_u, \sigma_u)$
    - $v_i \sim N(\mu_v, \sigma_v)$
- Hyper-priors control the population-level means and standard deviations:
    - For the means mu_u and mu_v, we use Gaussian priors:
        - $\mu_u \sim N(90,10)$
        - $\mu_v \sim N(5,2)$
    - For the standard deviations sigma_u and sigma_v, we use Half-Gaussian (HalfNormal) priors:
        - $\sigma_u \sim \text{HalfNormal}(10)$
        - $\sigma_v \sim \text{HalfNormal}(2)$

The HalfNormal priors ensure positive standard deviations, reflecting realistic variability in component behaviour.



This hierarchy structure allows partial sharing of information across components, critical when data is sparse.

We then applied a Bayesian algorithm (called Hamiltonian Monte Carlo) that repeatedly samples plausible values for all parameters (u_i, v_i, noise, weights (enhanced), etc.), based on the observed data. The result is not just a single number, but a full distribution of possible values, capturing the uncertainty in the predictions.

This HMC algorithm works by repeatedly proposing new plausible values for all the unknown parameters in our model. At each iteration of the sampling process, it then computes the probability of this new posterior based on the Bayesian equation:

$$\textbf{Posterior}(\theta' | \textbf{data}) \propto \textbf{Likelihood}(\textbf{data} | \theta') \times \textbf{Prior}(\theta')$$

θ being the set of parameters (…)

This equation tells us how plausible the new guess theta' is:

- The **Likelihood** measures how well the model with theta' would have explained the observed data.
- The **Prior** measures how reasonable theta' was considered to be before seeing the data.
- Based on the computed probability, the algorithm **decides whether to accept** the new proposal theta' (or stick with the previous one).

Each accepted sample adds to a growing set of examples of what the true parameters could be. Over many iterations, these samples build up a complete picture of the full posterior distribution.

1.2 Enhanced model

In the enhanced model, we again treat the degradation equations with uncertainty - but now we make the predictions even smarter by incorporating extra information about each component's characteristics (called diagnostics: X_1 to X_5).

This is the core of the enhanced predictive model.

- The observation model is still Gaussian, with observation noise sigma_y.
  - The noise sigma_y is shared across all components and has a half-Gaussian (HalfNormal) hyper-prior.
- The predictor now adjusts the basic exponential decay based on the component diagnostics:

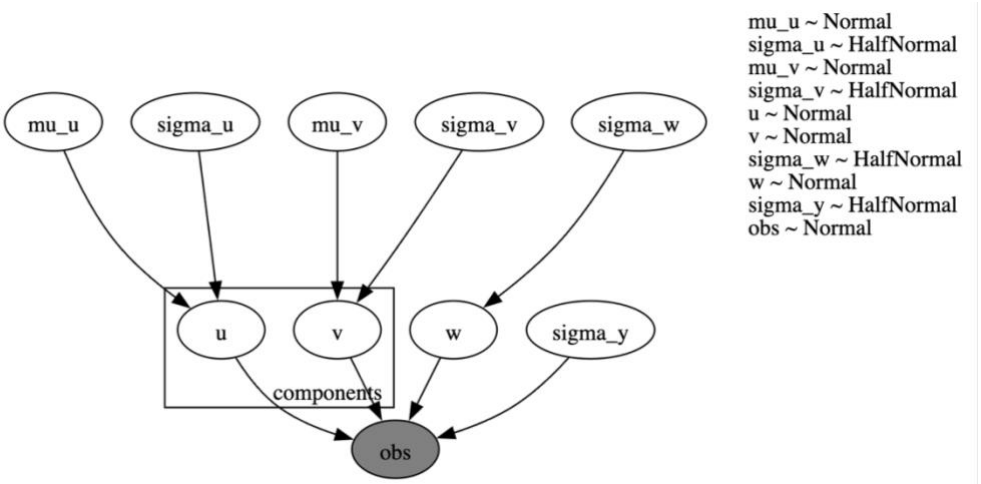
$$\mu(t) = u_i \exp\left(-\frac{(v_i + X_i \cdot w)t}{100}\right)$$

where:

w is a vector of learned weights describing how the diagnostic features (X_1–X_5) affect degradation.

- Priors over all u_i and v_i remain Gaussian:
  - ui∼N(μu,σu)
  - vi∼N(μv,σv)
- Hyper-priors again control the population-level means and standard deviations:
  - For the means mu_u and mu_v, we use Gaussian priors:
    - μu∼N(90,10)
    - μv∼N(5,2)
  - For the standard deviations sigma_u and sigma_v, we use Half-Gaussian (HalfNormal) priors:
    - σu∼HalfNormal(10)
    - σv∼HalfNormal(2)

- A **global weight vector** w is now introduced to model the influence of diagnostics:
  - Each w_j∼ N(0, sigma_w)
  - The overall variability sigma_w itself has a half-Gaussian prior:
    - sigma_w∼HalfNormal(1.0)

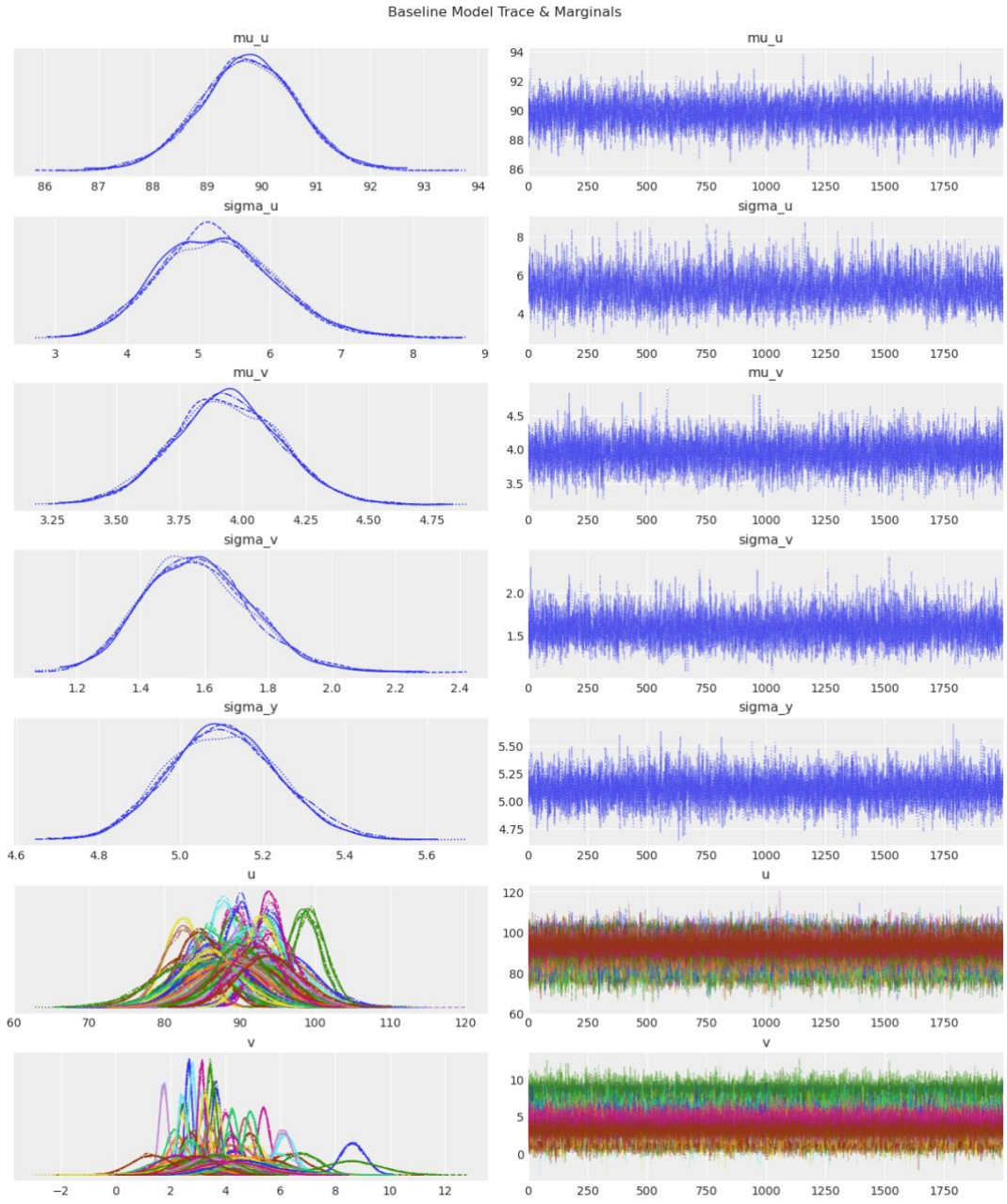


The key new addition is the global weights w, allowing diagnostic measurements to explain component-to-component variation.

This richer structure enables better predictions, especially when diagnostic data provides strong signals about future degradation. This gives a powerful predictive engine that doesn’t just make a "best guess," but captures the full uncertainty about component integrity, informed by both past observations and diagnostic features.

1.3 Diagnostics

Baseline model diagnostics:



All diagnostics showed excellent convergence; R^ values are 1.00, effective sample sizes are large, and no divergences were observed.

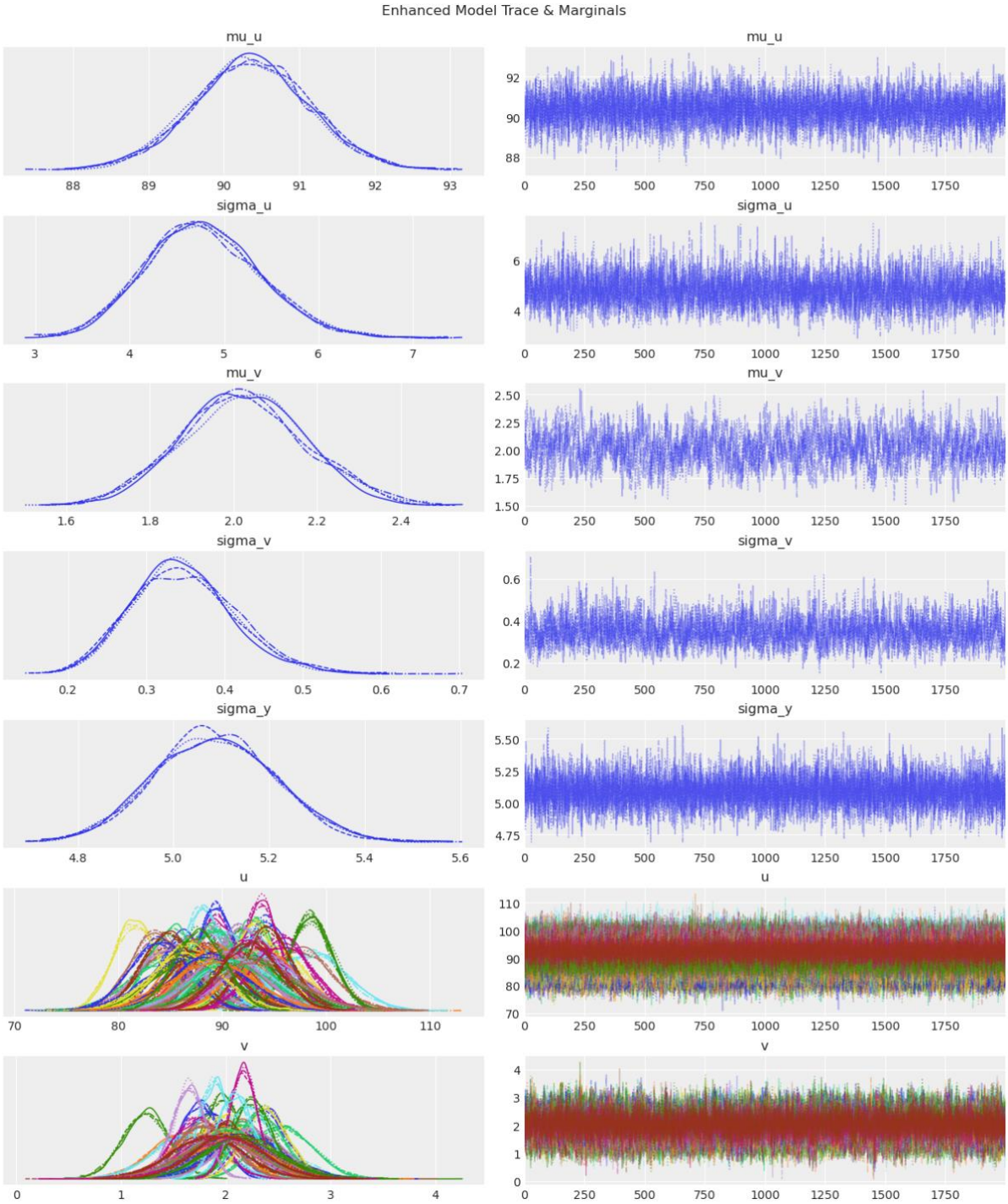
The baseline model's trace plots show that the sampling has performed well. The traces for key parameters like mu_u, sigma_u, mu_v, sigma_v, and sigma_y appear stable and densely populated, with no signs of divergence, drift, or poor mixing. This indicates that the chains have converged properly, and the sampler is effectively exploring the parameter space.

The marginal distributions are smooth and unimodal for all main parameters, suggesting that the model has identified clear, well-supported estimates for each. For example, mu_u is

centred around 90 (reflecting the typical starting integrity), and mu_v around 4 (capturing the average degradation rate).

The individual parameters u and v show reasonable spread across components, as seen in the colourful marginal plots at the bottom. This means the model is capturing real component-to-component variability without instability. Overall, the diagnostics show that the baseline Bayesian model has fitted the data successfully, producing reliable and interpretable parameter estimates.

Enhanced model diagnostics:

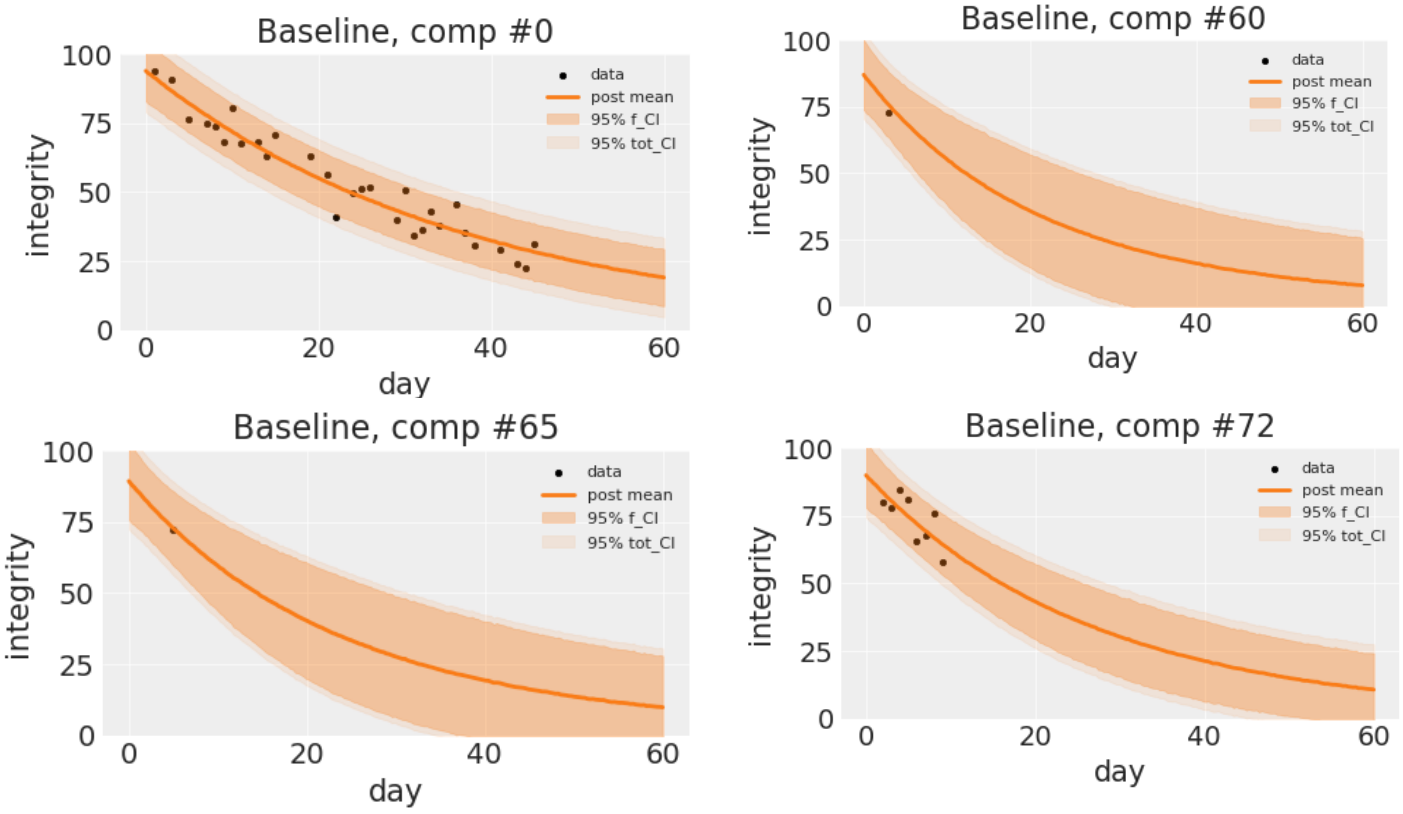


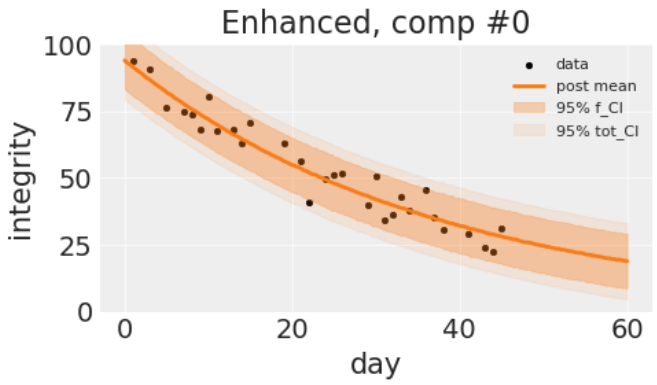
The trace plots for the enhanced model again show good sampling behaviour. The chains for parameters like mu_u, sigma_u, mu_v, sigma_v, and sigma_y are stable, well-mixed, and stationary, indicating strong convergence. The marginal distributions are smooth and single-peaked, confirming that the model finds clear estimates. The addition of diagnostic weights slightly increased model complexity, but the u and v parameters remain well-behaved across components. Overall, the enhanced model successfully captures the variation in the data and provides robust, interpretable posterior distributions.

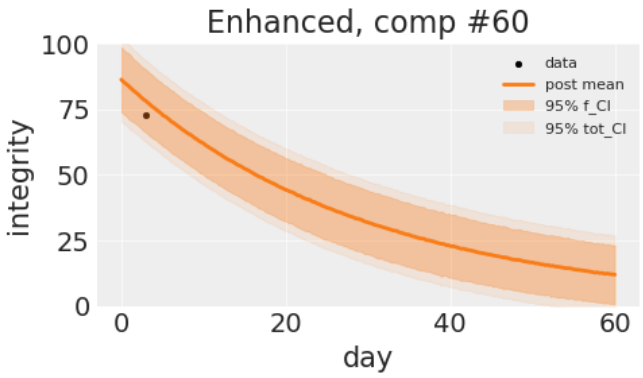
In the **enhanced model**, the variability in the degradation rates (v) is noticeably *lower* (see narrower spread in v's marginals), because the model explains some of the differences between components through the diagnostic weights. In contrast, the **baseline model** had to absorb *all* variation into v itself, since it did not have any X1–X5 information to help explain differences.

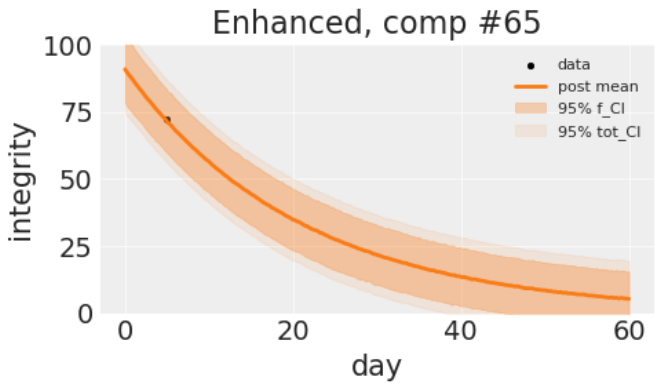
1.4 Illustrations of the Model Fit for Four Individual Components

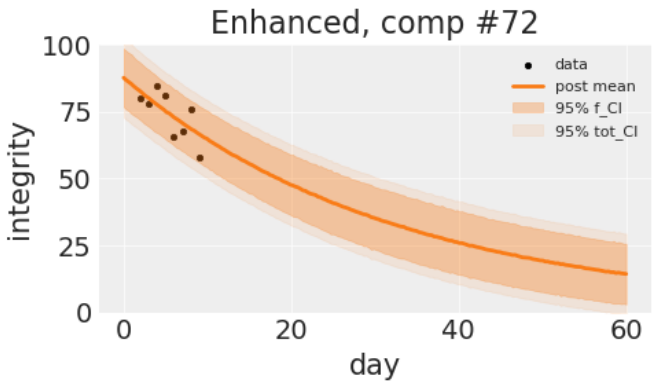
We illustrate model performance with four components; one well-observed component and **t**hree sparsely observed components.











Component #0 (Plentiful Data):

For component 0, where we have many observations across time, **both models** (baseline and enhanced) fit the data very well.

- The prediction line closely follows the data points.
- The uncertainty bands are tight, reflecting high confidence. This shows that when the model has enough data for a component, both approaches are capable of producing accurate predictions.

Components #60, #65, and #72 (Limited Data):

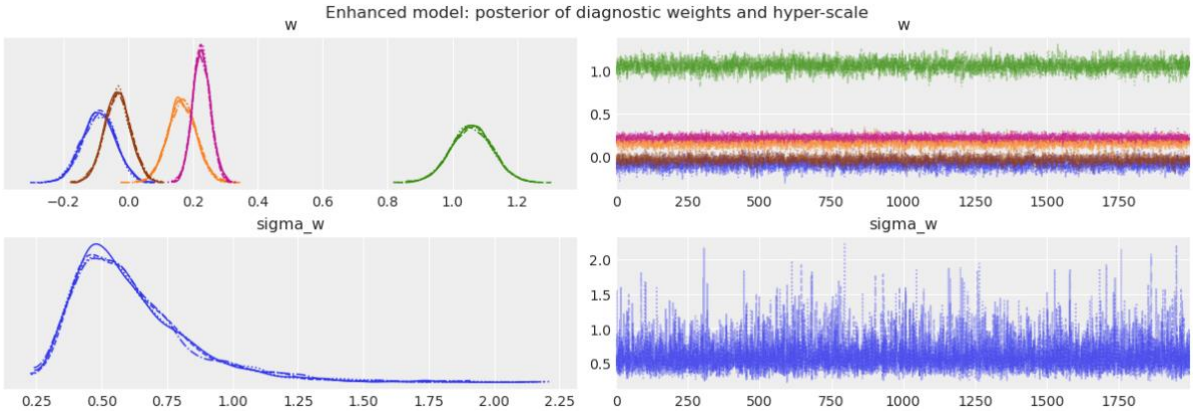
For the other three components, the number of observed data points is much smaller. Here, we see a more noticeable difference between the two models:

- In the **baseline model** plots, the prediction lines often over- or under-estimate the limited data, and the uncertainty intervals are wider. The model struggles slightly more because it does not incorporate additional information beyond the basic decay pattern.
- In the **enhanced model** plots, the prediction lines better track the limited data even with few points. The uncertainty intervals are slightly narrower. This is because the enhanced model can "borrow strength" from the additional features (X1–X5 diagnostics) to refine its prediction for each component.

Overall, the illustrations show that both models perform similarly when there is abundant data, but the enhanced model shows clear advantages when data is sparse. By using extra

diagnostic features, the enhanced model can make more informed predictions and reduce uncertainty even when only a few integrity measurements are available.

2 Potential Causes of Degradation



Using the output data and trace plots for the enhanced model, we can draw the following clear conclusions about the importance of each diagnostic feature:

- **Diagnostic 3 (X_3)**, corresponding to $w_3 \approx 1.06$, is by far the strongest and most important predictor of component degradation. Its 95% confidence interval does **not include zero**, providing strong evidence that X_3 has a real and meaningful impact on the decay rate.
- **Diagnostic 4 (X_4)** and **Diagnostic 2 (X_2)**, with weights around **0.23** and **0.17** respectively, also have positive and credible non-zero effects, although their impact is smaller than X_3.
- **Diagnostic 1 (X_1)** shows a small negative effect on the degradation rate—just barely credibly below zero—which suggests it might be mildly protective against wear.
- **Diagnostic 5 (X_5)** appears to have no meaningful effect; its 95% confidence interval spans both positive and negative values, indicating uncertainty and no reliable influence on degradation.

The hyper-parameter **sigma_w ≈ 0.62** reflects **moderate variability** across the effects of different diagnostics. However, most of the variability is explained by Diagnostics 2–4, with Diagnostics 1 and 5 playing minimal roles.

2.1 Practical Implications

Overall, **Diagnostic 3 (X_3) clearly stands out** as the most critical factor for predicting and understanding component degradation. A one-unit increase in X_3 leads to an estimated increase in the decay rate by about **1.06**, with very tight uncertainty bounds. Diagnostics X_2 and X_4 also contribute useful predictive information, though to a lesser extent.

From a practical standpoint, **focusing on monitoring and controlling the factor measured by X_3** would likely have the **biggest impact** on improving the forecasting of wear—and potentially extending the lifespan of the components. Meanwhile, X_2 and X_4 offer secondary benefits, while X_1 and X_5 may be less critical to prioritize.

3 Addressing team Queries

3.1 Black-box Model vs Bayesian Hierarchical Model

To illustrate why a Bayesian hierarchical approach is superior to a simple black-box model, I trained a "black-box-style" function to predict the integrity value at **t = 30**, based only on early observations from **t ∈ [0, 10] and the component diagnostics X1-X5**.

Model Description

For each component, I built a feature vector comprising:

1. Mean Integrity (0–10 days)**:** the average of all integrity measurements up to t = 10.
2. Average Decay Gradient**:** the mean of the finite differences (Δintegrity/Δtime) computed across its observations in [0, 10].
3. Diagnostics (X_1–X_5)**:** the five physical component characteristics provided by the client.

Model Training

Using the components with known integrity at day 30, we assembled these features to train a Random Forest regressor — a popular non-Bayesian machine learning method.
The Random Forest learns a flexible mapping:

$$[\text{mean\_intergity}, \text{avg\_gradient}, X1, \ldots, X5] \longrightarrow \text{Integrity at day 30}$$

The Random Forest model works by growing an ensemble of decision trees. Each tree provides a prediction, and the forest averages across them to make a final estimate. This technique is simple, fast, and captures complex, non-linear patterns without needing much manual tuning.

However, the black-box model has **important limitations**:

- It provides only a single point estimate for each component — with no measure of uncertainty.
- If early data is sparse, the Random Forest still "forces" a confident prediction without acknowledging that uncertainty should be higher.

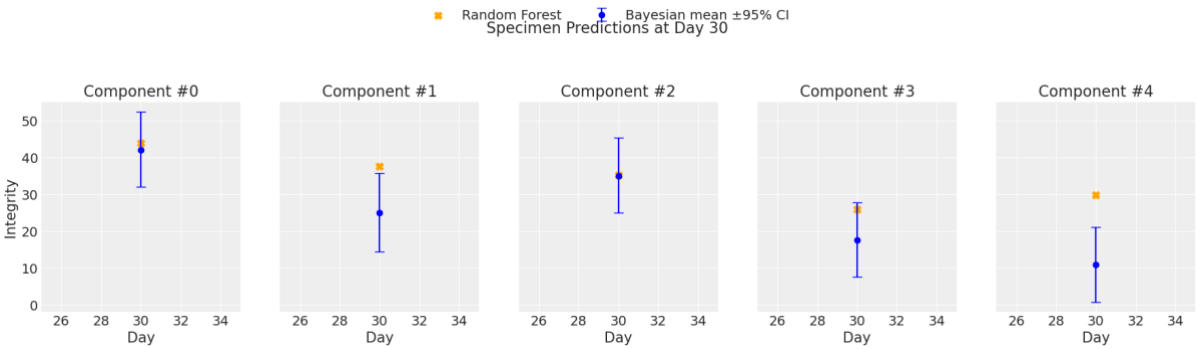
Why the Bayesian Approach is Better

In contrast, the Bayesian hierarchical model:

- Shares information across components automatically, allowing better predictions even with little early data.
- Produces a full probability distribution (credible intervals) for each prediction.
- Naturally shows wider uncertainty when less information is available, rather than overconfident guesses.

This makes the Bayesian model much better suited for real-world maintenance decisions, where knowing the confidence in a prediction is just as important as the prediction itself.

Specimen Prediction Graphs: Black-box vs Bayesian



The specimen prediction graphs show how the black-box Random Forest model and the Bayesian hierarchical model perform when predicting integrity at day 30.

- The orange X represents the Random Forest prediction — a single point with no information about uncertainty.
- The blue dot with error bars represents the Bayesian prediction — showing not only a best estimate (the mean) but also a 95% credible interval (the uncertainty range).

From the graphs we can see the Random Forest makes one fixed guess per component, regardless of how much early data was available. The Bayesian model provides a **range of plausible values** for each component. When early data is scarce or noisy, the Bayesian credible interval automatically becomes wider, reflecting greater uncertainty.

This visual contrast highlights a key advantage of the Bayesian approach: it does not just predict *what* the integrity might be, but also *how sure* we are. In operational settings - like deciding whether to refurbish a component - this uncertainty information is critical for making safer, more informed decisions.

3.2 The Support Vector Machine (2 Marks)

Support Vector Machines (SVMs) are powerful for clean, well-labelled data, but they are not ideal for our task of predicting component degradation under uncertainty. In this project, a hierarchical Bayesian model is better suited because:

- **Uncertainty matters:** We need not just a point prediction but also a full understanding of the uncertainty, which Bayesian methods naturally provide.
- **Sparse, uneven data:** Some components have little data. Hierarchical modelling shares information across components, whereas SVMs would likely overfit.
- **Interpretability:** Bayesian models allow us to clearly quantify how different diagnostics impact degradation, unlike the black-box nature of SVMs.
- **Integration of known physics:** Our model directly incorporates the known exponential decay behaviour, something SVMs cannot easily exploit.

Essentially, the Bayesian approach gives us more reliable predictions**,** better uncertainty estimates, and deeper insights — all critical for real-world engineering decisions.

4 Blind Test predictions

I have attached a text file “predictions.csv” containing 25 predictions for the probability that integrity is 30% or under at t=30 for components index 50-74.

Components with a high probability of falling below 30% integrity at day 30 are strong candidates for refurbishment or replacement. Specifically, components with probabilities above 50% are at significant risk and should be prioritized. Components with moderate probabilities (around 20–50%) may also warrant closer monitoring or preventative maintenance. On the other hand, components with very low predicted probabilities (below 10%) are expected to remain in good condition and can safely continue in service without immediate intervention.