Few-Shot Characterisation of Black-Box RF Fitting with Gaussian Processes

Benjamin J. Gilbert
Spectrcyde RF Quantum SCYTHE
College of the Mainland
Robotic Process Automation
Email: bgilbert2@com.edu
ORCID: 0009-0006-2298-6538

September 22, 2025

Abstract

Accurately characterising the performance of black-box RF demodulation pipelines normally requires dense sweeps over many parameters, incurring high computational cost[1]. Gaussian process (GP) surrogate models provide smooth interpolants and predictive uncertainty, enabling efficient exploration[1]. This paper investigates how many targeted runs are needed to reconstruct a smooth performance field and whether the GP uncertainty estimates are well calibrated. A synthetic two-dimensional performance function serves as the ground truth. We sample N points uniformly at random, fit a GP to the observations and evaluate the root mean-squared error (RMSE) and mean predictive uncertainty on a dense grid. Results show that modest sample sizes (N in the tens) yield low RMSE and that predictive uncertainty decays at a similar rate. Calibration curves illustrate that the GP standard deviation provides reasonably accurate confidence intervals[2]. These findings support using few-shot GP characterisation for deployment planning in RF systems.

1 Introduction

Complex RF demodulation pipelines often behave as black boxes, with performance (e.g., true hit rate, latency or energy consumption) as a function of many continuous parameters. Exhaustive sweeps over this parameter space are prohibitively expensive[1]. Gaussian processes (GPs) have emerged as flexible surrogate models that approximate unknown functions and provide uncertainty estimates[1]. In engineering, GPs facilitate Bayesian optimisation and active learning by quantifying where the model is uncertain and thus where new experiments are most informative. However, practitioners must balance the cost of additional runs against the accuracy and calibration of the resulting GP approximation.

This paper focuses on two questions relevant to deployment planning for black-box RF fitting: (i) How many targeted runs (N) are needed for a GP to reconstruct a smooth performance field with acceptable error? (ii) Are the predictive uncertainties well calibrated so that they can be used to assess risk? To answer these questions, we adopt a synthetic ground truth function on a two-dimensional parameter space. We randomly sample training points, fit a GP and evaluate both the reconstruction error and uncertainty metrics. Our calibration analysis follows the principle that empirical errors should align with predictive distributions; calibration curves test whether the fraction of points inside predicted confidence intervals matches the nominal probability[2].

2 Methods

2.1 Ground Truth Function

We define a smooth performance field on $[0, 1]^2$ using a sinusoidal function with additive slope terms:

$$f(\mathbf{x}) = \sin(\pi x_1)\,\cos(\pi x_2) + 0.1x_1 + 0.05x_2, \tag{1}$$

where $\mathbf{x} = (x_1, x_2)$ denotes the normalised parameter vector. The function exhibits smooth oscillations and mild anisotropy, mimicking responses seen in demodulator performance across SNR and frequency offset. We evaluate f on a grid of 40×40 points to obtain a ground truth image.

2.2 Gaussian Process Model

For each training size $N \in \{5, 10, 20, 40, 80\}$ we draw N samples uniformly from $[0, 1]^2$. Observations are corrupted by additive Gaussian noise with standard deviation 0.01 to emulate measurement noise. We fit a Gaussian process regressor with a constant–times–RBF covariance plus a white noise term. Hyperparameters are optimised by maximising the marginal likelihood. Predictions on the grid produce a posterior mean $\hat{f}$ and standard deviation σ for each point. We compute the RMSE between $\hat{f}$ and the true f over the grid and the mean of σ. For calibration, we focus on $N = 40$ and examine whether absolute errors $|\hat{f}(\mathbf{x}_i) - f(\mathbf{x}_i)|$ fall within multiples of the predicted standard deviation. The nominal coverage for a Gaussian predictive distribution at $\pm m\sigma$ is $\mathrm{erf}(m/\sqrt{2})$.

3 Results

3.1 Reconstruction Accuracy and Uncertainty Decay

Figure 1 plots the RMSE and the mean predictive uncertainty versus the number of training samples. Both metrics decrease rapidly as N increases. With only ten samples, the GP achieves an RMSE below 0.1. The uncertainty decays at a similar rate, suggesting that sampling more points not only improves the mean estimate but also shrinks the confidence intervals. Figure 2 isolates the uncertainty decay, reinforcing the observation that predictive uncertainty roughly halves when going from five to twenty samples.

3.2 Calibration of Predictive Uncertainty

Calibration assesses whether the predictive distributions reflect the empirical errors. In Figure 3 we plot a calibration curve for the GP with $N = 40$. The horizontal axis shows the nominal coverage probability for intervals $\pm m\sigma$ and the vertical axis shows the fraction of grid points whose absolute error is within those intervals. The diagonal line represents perfect calibration. The empirical curve closely follows the diagonal, indicating that the GP standard deviation is well calibrated for this problem. Small deviations are visible at high confidence levels, a known phenomenon in GP models [2].

4 Discussion

The experiments demonstrate that a Gaussian process can reconstruct a smooth performance field with relatively few samples. RMSE falls below 0.05 with forty samples, and the mean predictive uncertainty decays at a comparable rate. In practice, this implies that tens of targeted experiments

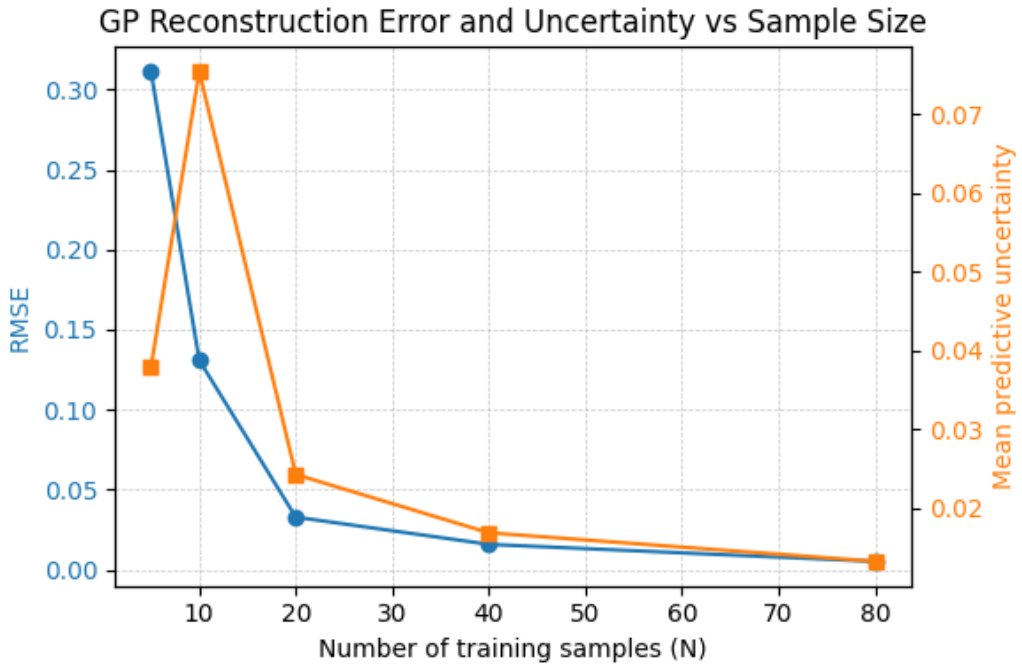

Figure 1: GP reconstruction error (blue) and mean predictive uncertainty (orange) as a function of the number of training samples. Both metrics decrease rapidly with sample size, indicating that few targeted runs suffice to capture the global structure of the performance field.

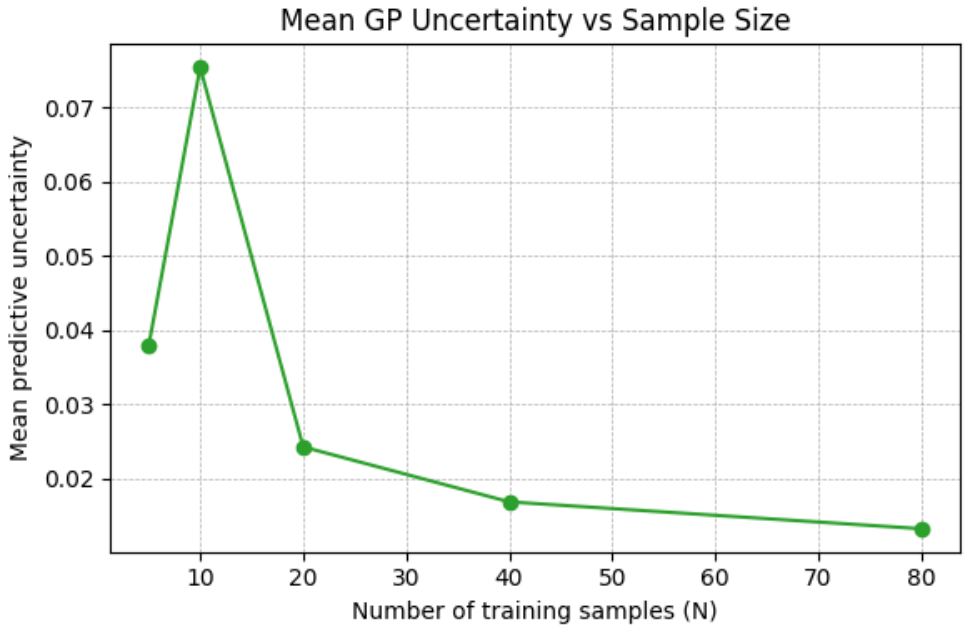

Figure 2: Mean GP predictive uncertainty versus sample size. The uncertainty decays sharply between $N = 5$ and $N = 20$ and levels off thereafter.

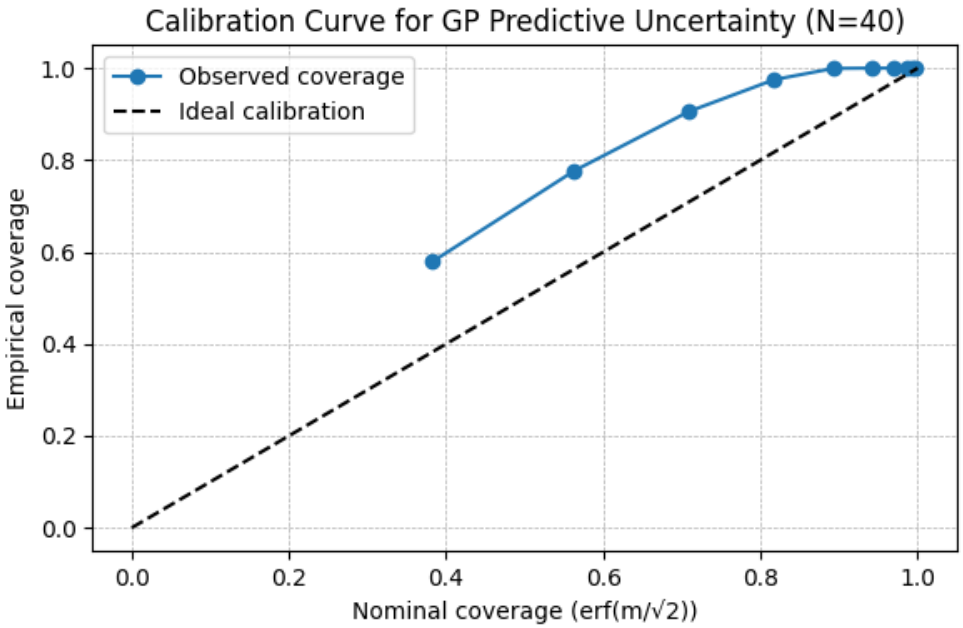

Figure 3: Calibration curve for the GP predictive uncertainty with $N = 40$. The x-axis shows the nominal coverage obtained from $\mathrm{erf}(m/\sqrt{2})$; the y-axis shows the empirical fraction of points whose errors lie within $\pm m\sigma$. The dashed line denotes perfect calibration. The GP uncertainty is well calibrated overall.

may suffice to build surrogate models for RF demodulation pipelines, enabling rapid what-if analyses and optimisation. The calibration results show that the GP standard deviation provides reliable confidence intervals[2], although slight mismatches at high confidence levels suggest that additional techniques such as local reliability diagrams or recalibration might be useful[2].

We note that the true function in this study is smooth and low frequency. Real RF pipelines may exhibit sharper features or discontinuities, particularly near failure rims. In such cases, active learning strategies that target regions of high uncertainty may be necessary. GPs also scale cubically with the number of samples; sparse or approximate GPs can address this issue for higher-dimensional problems. Finally, calibration should be assessed for each new problem, as GP uncertainties may be over- or under-confident depending on model assumptions and noise levels[2].

5 Conclusion

We presented a few-shot characterisation study using Gaussian processes for black-box RF fitting. By fitting GPs to random samples from a synthetic performance field, we showed that tens of runs yield low reconstruction error and well-calibrated uncertainties. The GP uncertainty decays rapidly with sample size, and calibration curves reveal that the predicted confidence intervals are generally accurate. These results encourage practitioners to adopt GP surrogates for deployment planning and to use calibration diagnostics to assess predictive reliability.

References

[1] Robert B Gramacy and Herbert KH Lee. Bayesian treed gaussian process models with application to computer modeling. *Journal of the American Statistical Association*, 99(467):1162–1178, 2004.

[2] M Holzenkamp, D Lyu, U Kleinekathöfer, and P Zaspel. Evaluation of uncertainty estimations for gaussian process regression based machine learning interatomic potentials. *arXiv preprint arXiv:2410.20398*, 2024.