Summer 8-6-2024

Uncertainty Quantification in Machine Learning Models Via Gaussian Process Regression: A Comparative Study

Ayorinde E. Olatunde
Case Western Reserve University, aeo49@case.edu

Weiqi Yue
Case Western Reserve University, wxy215@case.edu

Pawan K. Tripathi
Case Western Reserve University, pkt19@case.edu

Roger H. French
Case Western Reserve University, rxf131@case.edu

Anirban Mondal
Case Western Reserve University, axm912@case.edu

Author(s) ORCID Identifier:

Ayorinde Emmanuel Olatunde

Materials Data Science for Stockpile Stewardship

COE: US-Department of Energy-NNSA Award

Uncertainty Quantification in Machine Learning Models Via Gaussian Process Regression: A Comparative Study

Ayorinde E. Olatunde[1,4], Weiqi Yue[3,4], Pawan K. Tripathi[2,4], Roger H. French[2,3,5], Anirban Mondal[1,4,*]

[1]Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University , Cleveland, OH, 44106, USA

[2]Department of Materials Science and Engineering, Case Western Reserve University , Cleveland, OH, 44106, USA

[3]Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA

[4]Materials Data Science for Stockpile Stewardship: Center of Excellence, Case Western Reserve University, Cleveland, OH, 44106, USA

*axm912@case.edu

1. INTRODUCTION

Background of the Study:

- Synchrotron X-ray Diffraction (SXRD) problems are solved using Machine Learning (ML). Uncertainty Quantification (UQ) ensures model reliability and trust.

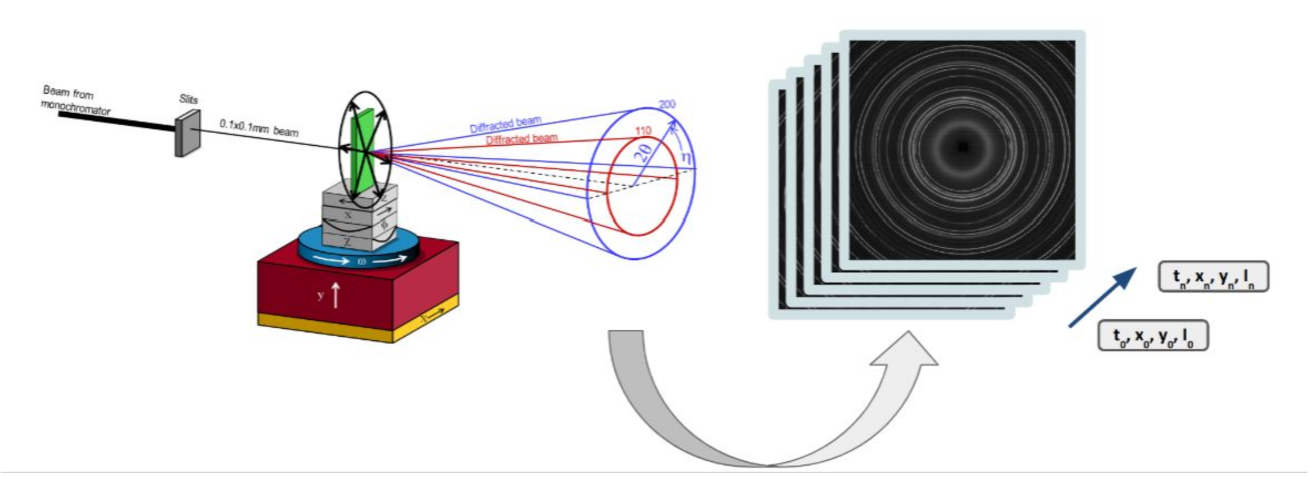
Motivation of the Study:

- QU is harder feature space increases[1]. Additive covariance kernels help with UQ in large feature spaces[2]

Goal of the Study:

- Extend UQ for predicting β-phase volume fraction in Ti-6Al-4V alloy[3] alloy using 2D diffraction images via Gaussian Process Regression (GPR) to higher feature spaces.

2. EXPERIMENTAL SET UP FOR DATA COLLECTION



The experiment involved generating time-series SXRD diffraction patterns of the sample by directing a beam of X-rays onto it as it underwent heating and cooling.

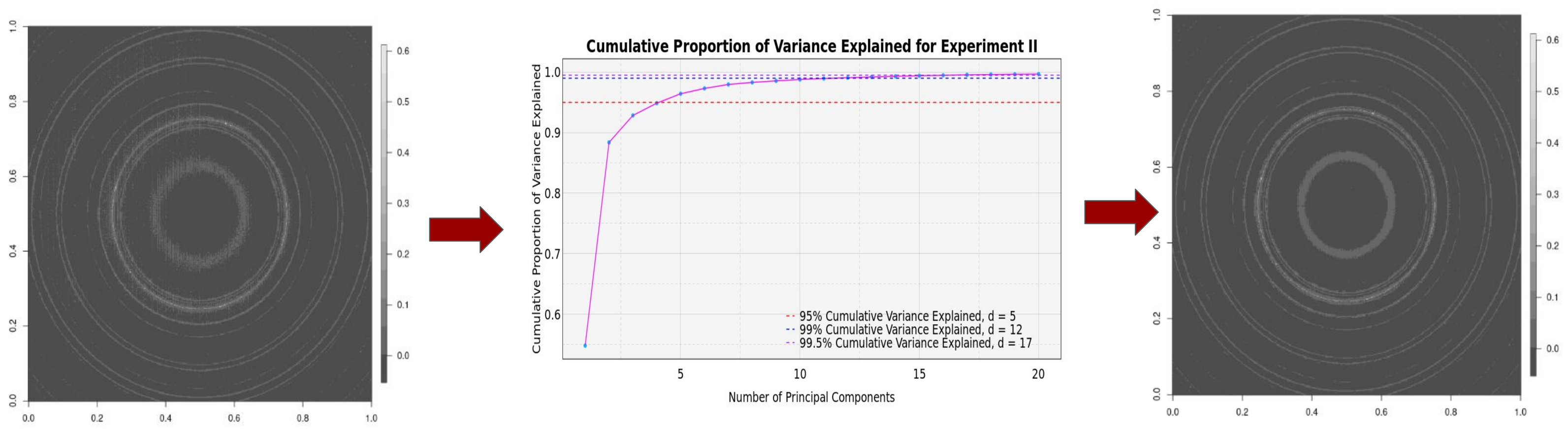
3. METHODOLOGY

Experimentation

Out of four experiments conducted, we used data set from three of them and a combination of the three data sets

Data Ingestion & Reconstruction

Apply PCA as dimension reduction technique for computational efficiency



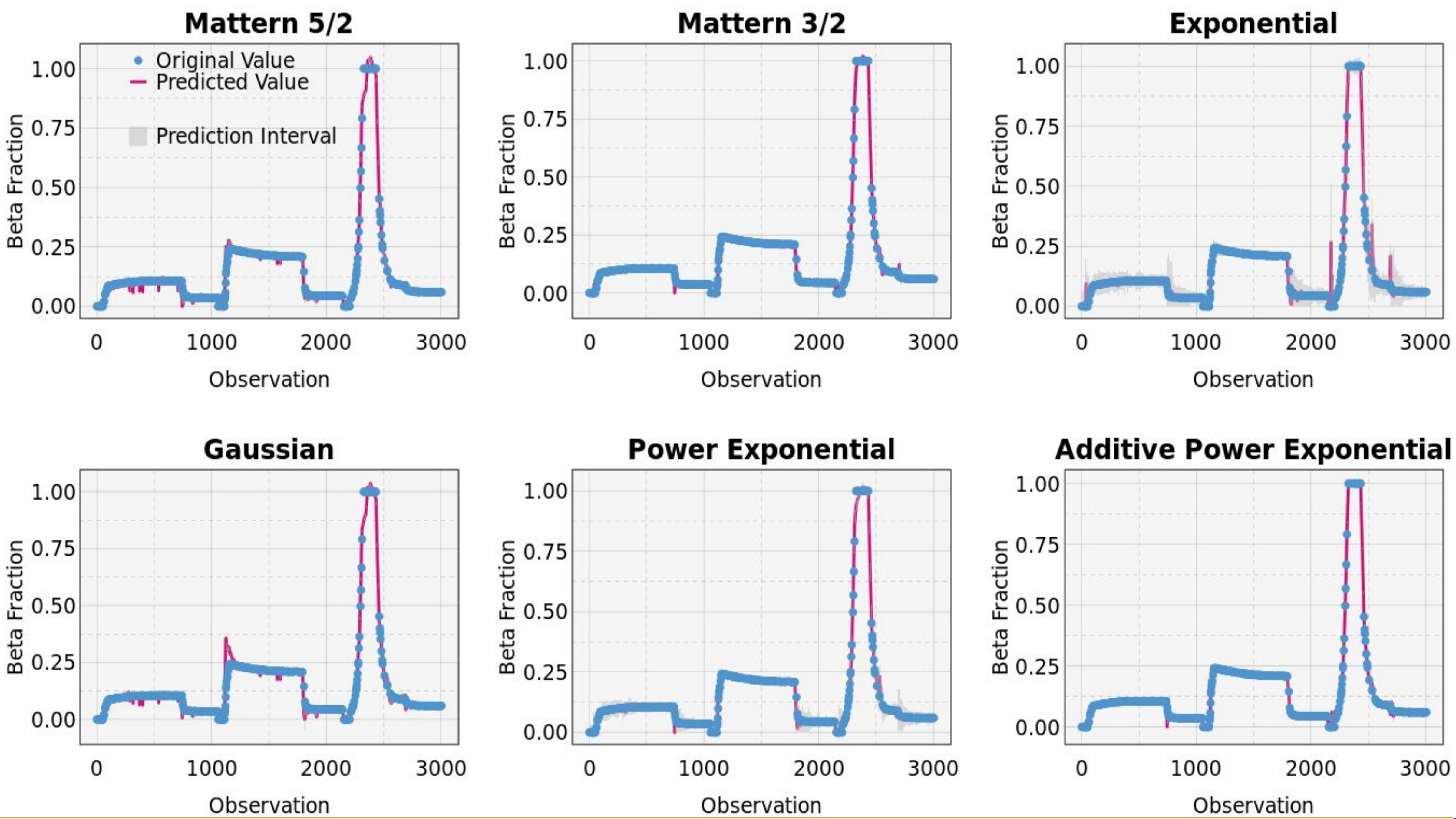
2048 X 2048

An example of reconstructed 2D diffraction image using 5 PC

Modelling with comparative Kernels & Quantifying Uncertainties

Used GPR, leveraging its mean and covariance functions for probability distribution over functional relationships, for predictions and UQ.

4. RESULTS



Original vs. Predicted β-fraction with 95% prediction interval (22 PCs) for test set of Combined data set. Each plot represents the model produced by the respective kernel structure.

- **Used six kernels for datasets from experiments I, II, III, and their combination, with sample sizes 1056, 1097, 855, and 3008, respectively**
- **Performed computation on HPC using 10 cores for I, II, III, and 30 cores for the combined dataset**
- **The plot and table show the results for the combined data set**

Model (Kernel)	Train (Time)	Train (RMSE)	Test (RMSE)
Mattern 5/2	1 hr 8 min 21 sec	0.0218	0.0201
Mattern 3/2	1 hr 33 min 16 sec	0.0070	0.0103
Exponential	1 hr 30 min 20 sec	2.8975e-05	0.0195
Gaussian	1 hr 7 min 48 sec	0.0292	0.0293
Power Exp	1 hr 9 min 18 sec	0.0121	0.0120
Add Pow Exp	4 min 57 sec	0.0013	0.0077

5. CONCLUSIONS & FUTURE DIRECTION

Conclusion

- **Extended methodology for UQ in ML models with higher feature space and reduced computation time**
- **Additive Exponential Kernels offer shorter computation times with similar UQ capabilities**

Future Direction

- **Expand work to handle larger sample sizes and integrate methods for managing both high sample sizes and high dimensions**

6. REFERENCES

[1] Bui-Thanh, Tan, et al. "Extreme-scale UQ for Bayesian inverse problems governed by PDEs." *SC'12: Proceedings of the international conference on high performance computing, networking, storage and analysis*. IEEE, 2012.

[2] Durrande, Nicolas, David Ginsbourger, and Olivier Roustant. "Additive covariance kernels for high-dimensional Gaussian process modeling." *Annales de la Faculté des sciences de Toulouse: Mathématiques*. Vol. 21. No. 3. 2012.

[3] Brown, D. W., *et al.* Evolution of the microstructure of laser powder bed fusion Ti-6Al-4V during post-build heat treatment. *Metallurgical and Materials Transactions* A 52 (2021): 5165-5181

7. ACKNOWLEDGEMENT

This material is based upon research in the **Materials Data Science for Stockpile Stewardship Center of Excellence (MDS3-COE)**, and supported by the **U.S. Department of Energy's National Nuclear Security Administration** under Award Number(s) **DE-NA0004104**. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at **Case Western Reserve University.**