SAND2022-12327C

A Functional Statistical Modeling Approach to Using Plutonium Particle Features from SEM Images

Pu Futures—The Science 2022

September 29, 2022

A. Zhang[1], A. McCombs[1], M. Ausdemore[2], D. Ries[1], J.D. Tucker[1], J.G. Huerta[1], K. Goode[1], and K. Shuler[1]

Department of Statistical Sciences

[1]Sandia National Laboratories

[2]Los Alamos National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Acknowledgements

We would like to thank our Pu Signatures team, consisting of colleagues at PNNL who implemented experimentation and sent us gigs of data, our SNL and LANL statistics colleagues and reviewers for the modeling work, experts at SRNL, ORNL, LANL, and LLNL for verifying our work, and our sponsors at NNSA and DHS for funding this work.

Problem Description

Goal of Nuclear Forensics: Identify and attribute a set of processing conditions to interdicted special nuclear materials.

- Knowledge about processing conditions is helpful in determining where the material originated
- Identifying processing conditions can be considered an **Inverse Prediction** (IP) problem
  - In classical regression, the covariates/predictors are known, and a statistical model is constructed to estimate their relationship with the response.
  - In IP problems, the goal is to estimate the covariates (processing conditions) using the observed responses.
  - Addressed using a **Functional Inverse Prediction** (FIP) framework from model incorporating information on particle features from SEM imagery.

Data

SEM images of particles were produced at different processing conditions levels specified from an I-optimal statistically designed experiment [1].

- The images were segmented and pre-processed to remove noise and background.
- MAMA morphological software was used to collect scalar features (measures of area, perimeter, major ellipse/minor ellipse, aspect ratio, etc.) for each particle and the response by run was represented by CDFs of the scalar features [2][3].
- Particle features are influenced by the processing conditions under which the material was produced.

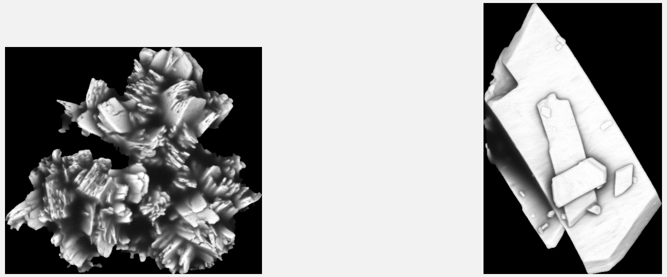


Fig.1: Contrasting particle features from different processing conditions.

Distributions of MAMA Measurements

X-axis: MAMA Characteristic
Y-axis: Cumulative Density Function (CDF)

Blue shading: True value of Condition 1 processing condition across *all* runs
Red line: True distribution of particles from a *single* run

These distributions can be treated as functions and modeled to discriminate between processing conditions.

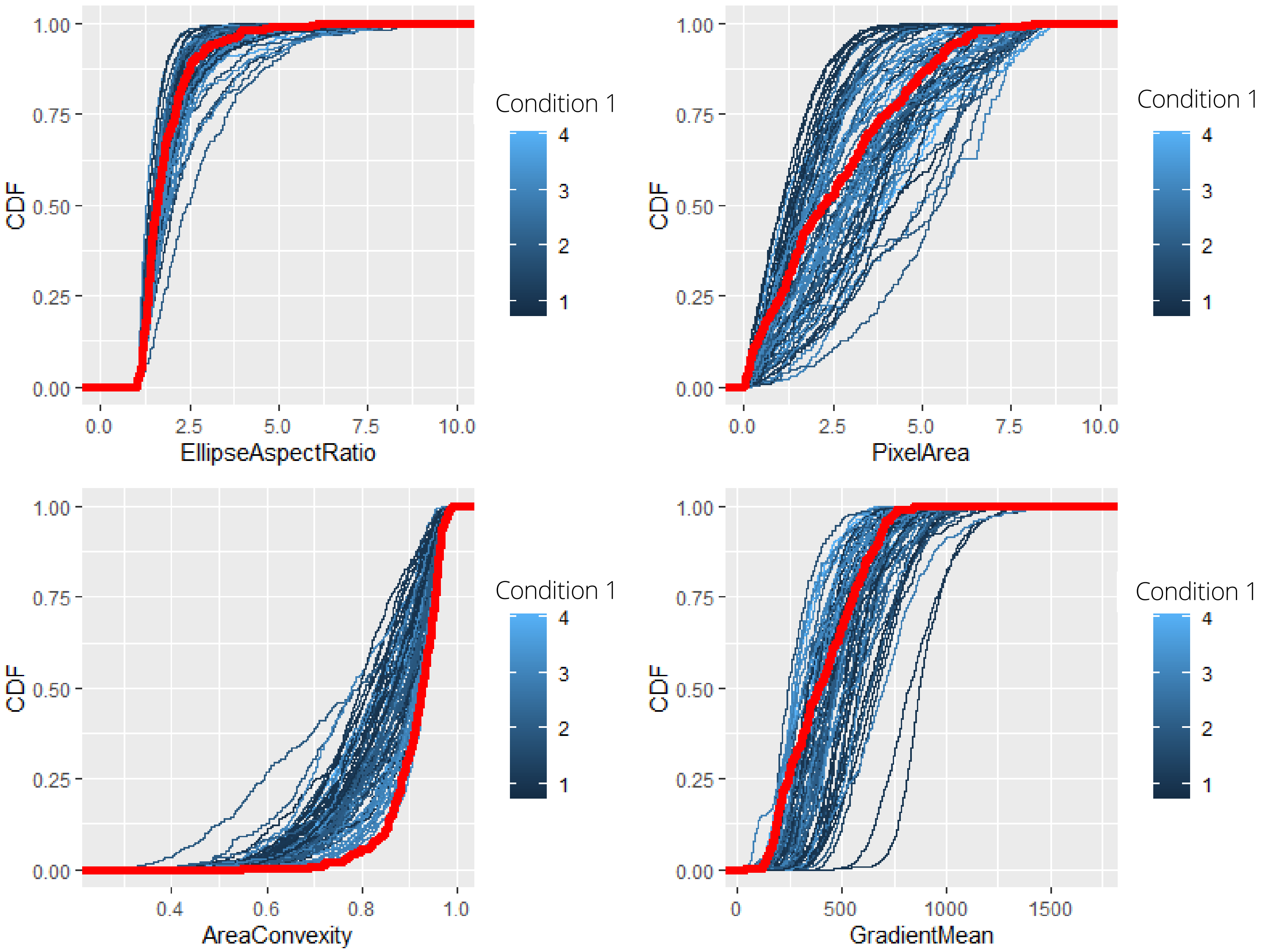


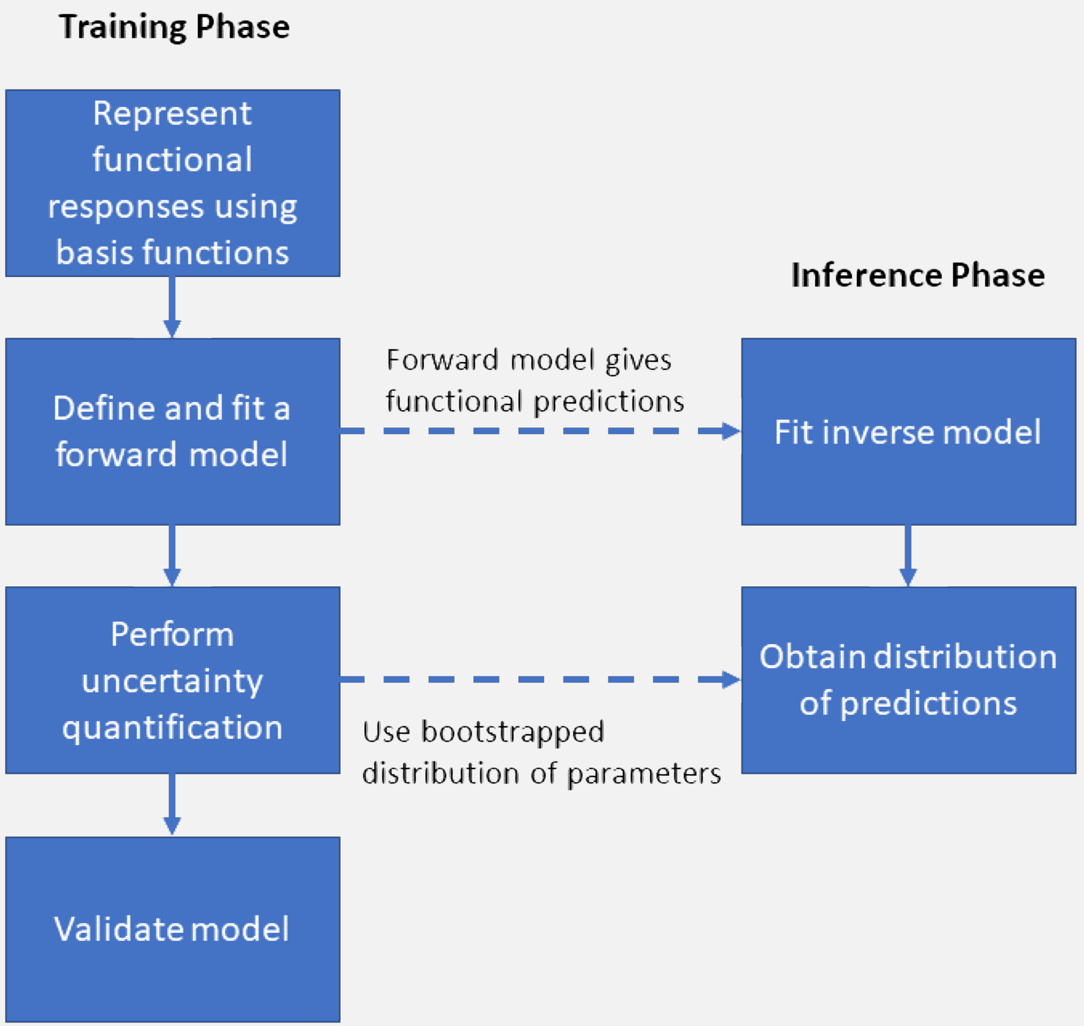
Fig. 2: Demonstration of representing CDFs of MAMA characteristics as functions.

Inference

The Functional Inverse Prediction (FIP) framework [4] is used to estimate material processing conditions, which has two stages:

1. Forward model is fit to training data
2. Relationships found from forward model are used to estimate unknown covariates from new responses

Fig. 3: Functional Inverse Prediction Framework.



Model

Forward Model:

$$Y_{ij}(t) = g_j\left(x_i; \theta_j(t)\right) + \epsilon_{ij}(t); i = 1, \dots, n; j = 1, \dots, q, \text{ where}$$

- $g_j(\cdot)$ can be taken to be $x_i' \beta_j(t)$ for functional linear regression, or other forms for non-linear responses
- x_i are processing conditions and $\epsilon_{ij}(t)$ are i.i.d., mean-zero, second-order stational stochastic processes

Inverse Model:

$$\hat{x}^* = \underset{X^*}{\operatorname{argmin}} \sum_{j=1}^{q} \int L(\hat{y}_j(t), Y_j^*(t))dt, \text{ where}$$

- $\hat{y}_j(t) = g_j\left(X^*; \hat{\theta}_j(t)\right), L(\cdot)$ is some loss function, Y_j^* are some new observations/functions.

Simulation Study

A simulated functional dataset was produced with two covariates,

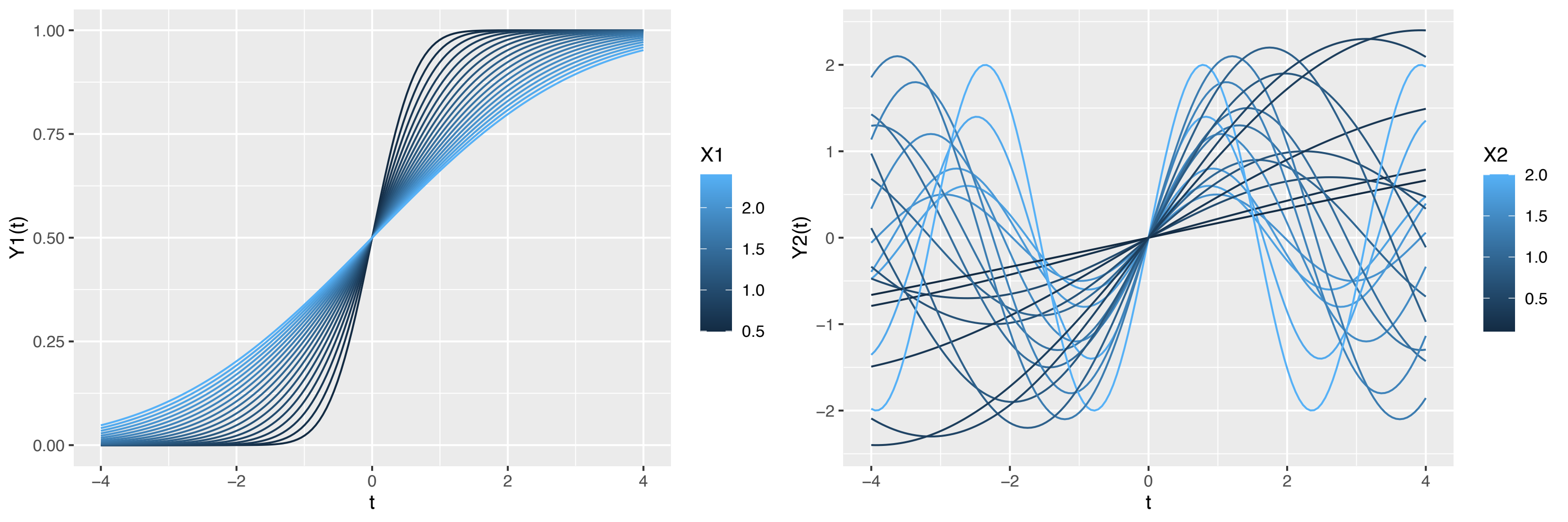


Fig. 4: Functional responses, Y1 and Y2, from simulated dataset.

$$Y_{i1}(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi x_{i1}}} e^{-\frac{1}{2x_{i1}^2}u^2} du \qquad\qquad Y_{i2}(t) = x_{i1}\sin(x_{i2}t)$$

- Functional linear model and MARS forward model fit using basis representation.

Simulation Study Results

The simulation study results indicate that the FIP framework is able to recover x_1 and x_2

Tab. 1: RMSE for IP of simulation study

	x_1	x_2
Linear	0.06 (0.04)	0.57 (0.26)
MARS	0.02 (0.04)	0.51 (0.26)

RMSE = standard deviation of the residuals

Approaches for FIP

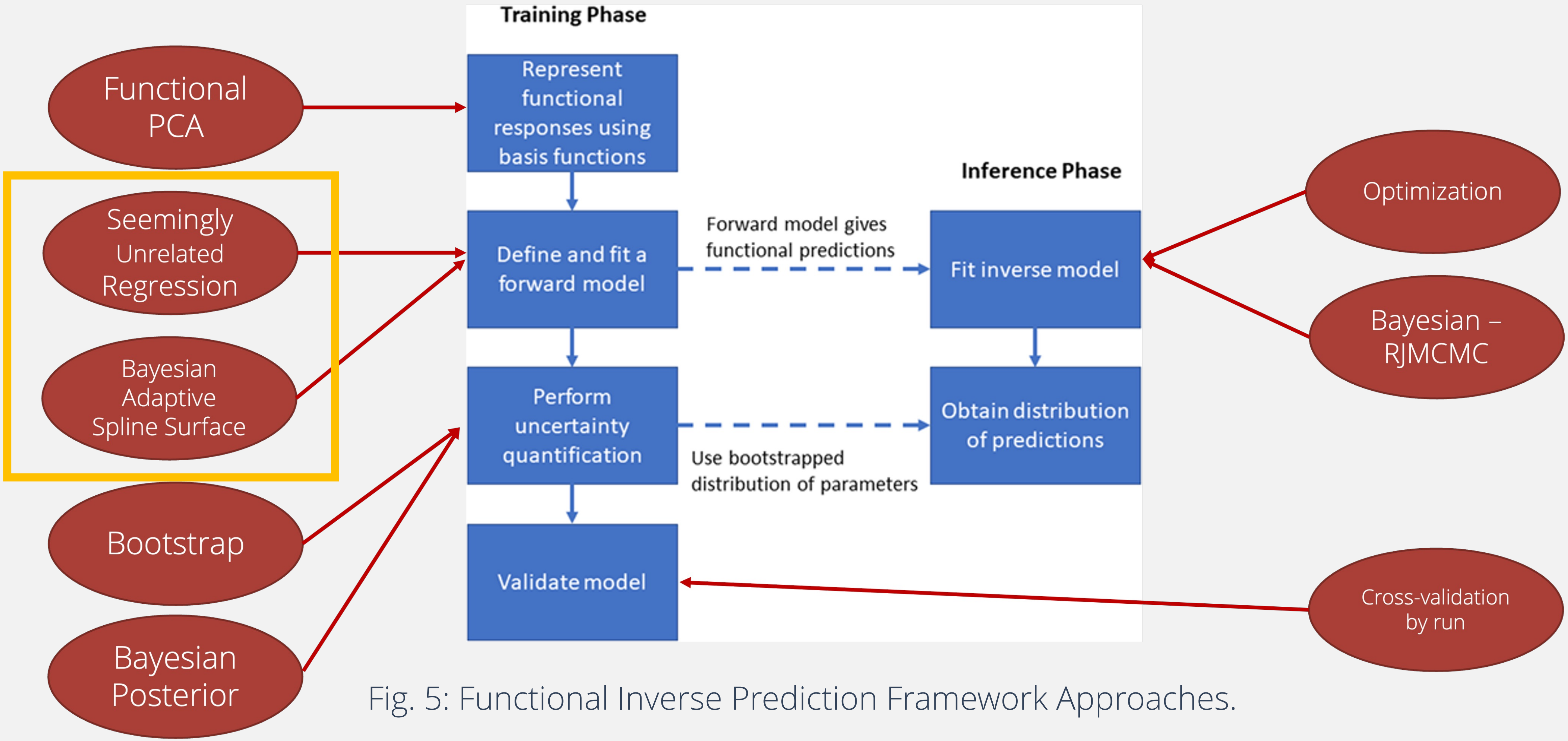


Fig. 5: Functional Inverse Prediction Framework Approaches.

Seemingly Unrelated Regression

- A generalization of a linear regression model that contains only exogenous regressors (only related through the error terms).
- Fit forward and inverse model
- Fit into a Bayesian framework for easy UQ
- Considers particles within a run as a group
- Accounts for correlations between errors in morphological characteristics. I.e. for a particular run with the same set of processing conditions, morphological characteristics, such as particle size, shape, and smoothness are likely related.

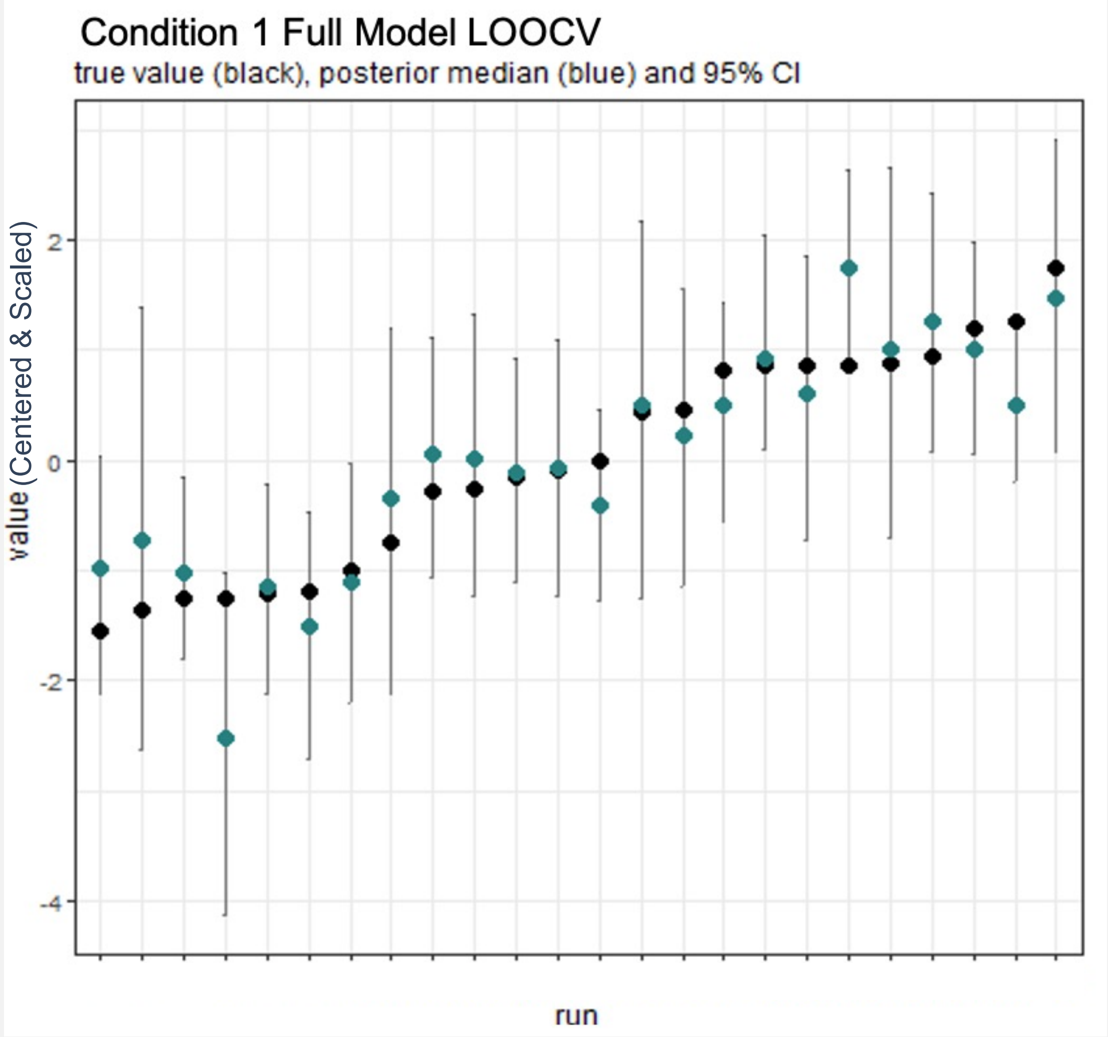
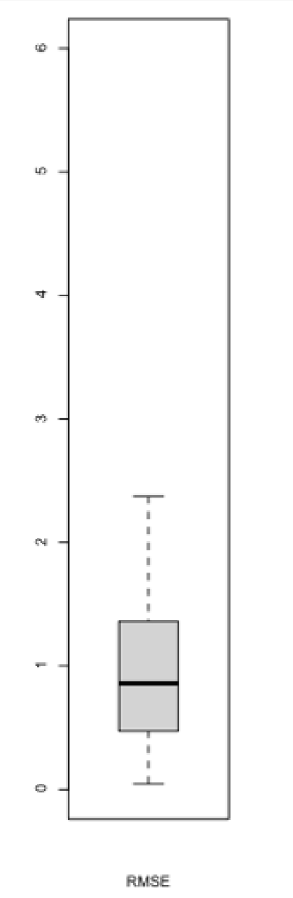


Fig. 6: Model performance assessment using LOOCV on SUR for Condition 1.

Bayesian Adaptive Spline Surface

- A non-linear model for nonparametric regression that incorporates flexibility, scalability, interpretability, and probabilistic accuracy.
- Fit a direct model
- Fits in a Bayesian framework for easy UQ
- Considers particles within a run individually, then combines results post hoc.

Fig. 7: BASS RMSE of Condition 1 for all particles



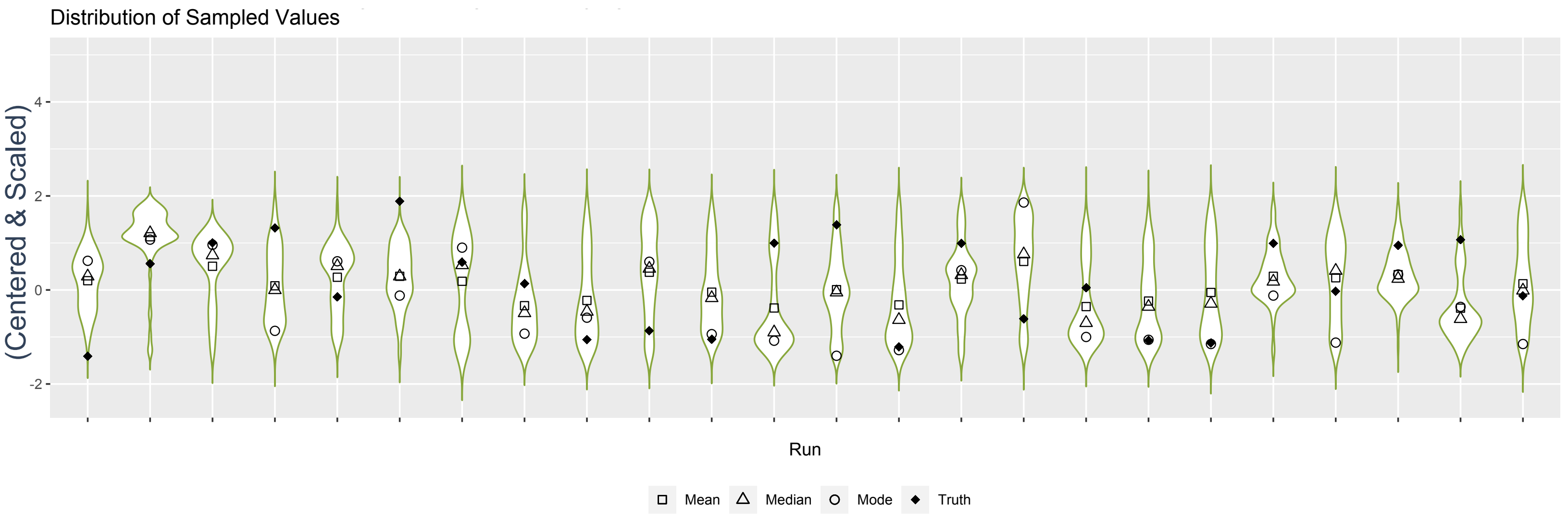


Fig. 8: Model performance assessment using LOOCV on BASS for Condition 1.

Conclusion

- There is a lot of variability in the chemical process that it makes it difficult to parse out the signal from noise in a purely-data driven manner.
- Non-linear relationships exist between processing conditions and particle features.
- Models are constructed so they are interpretable, allowing us to understand which particle features affect which processing conditions.
- Models are only as good as the data– if the major sources of variability are not captured in the data, then it will be difficult to develop generalizable data-driven models.
- Uncertainties on predictions are relatively large for all models considered.

Thank you!

Pu Futures attendees from statistics team:

Adah Zhang

azhang@sandia.gov

Madeline Ausdemore

mausdemore@lanl.gov

References

[1] C. Anderson-Cook, T. Burr, M. Hamada, C. Ruggiero and E. Thomas, "Design of experiments and data analysis challenges in calibration for forensics applications," *Chemometrics and Intelligent Laboratory Systems,* vol. 149, pp. 107-117, 2015

[2] Ries, D., Lewis, J., Zhang, A., et al. (2019). Utilizing Distributional Measurements of Material Characteristics from SEM Images for Inverse Prediction. *Journal of Nuclear Materials Management*.

[3] R. Porter, C. Ruggiero, N. Harvey, P. Kelly, L. Tandon, M. Wilkerson, K. Kuhn and D. Schwartz, "MAMA User Guide v 1.2. Technical Report," Los Alamos National Laboratory, Los Alamos, 2016.

[4] Ries, D., Zhang, A., Derek Tucker, J., Shuler, K., & Ausdemore, M. (2022). A framework for inverse prediction using functional response data. *Journal of Computing and Information Science in Engineering*, *23*(1), 011002.

[5] A. Zhang, C. Anderson-Cook, D. Ries, J. D. Tucker, K. Shuler, M. Ausdemore, M. Goldberg, D. Meier and R. Clark, "Statistical Analysis for Signatures of Plutonium (III) Oxalate Experiment," SAND2021-7765, Sandia National Laboratories, Albuquerque, 2021.