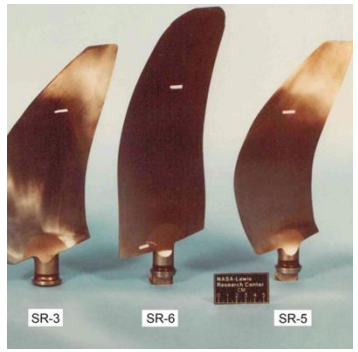
Glenn Research Center



Machine Learning Surrogates and Bayesian Optimization in Turbomachinery Structural Design

Kristopher Pierson
NASA Glenn Research Center

Advanced Modeling & Simulation (AMS) Seminar Series
NASA Ames Research Center, September 18th, 2025







This work was supported by:
NASA AATT Project
AATT: Advanced Air Transport Technology

Acknowledgements

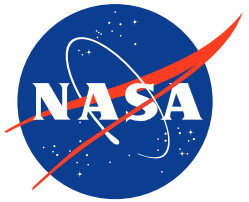
- John Gillespie
  - NASA Glenn Research Center
- Matthew Ha
  - HX5, LLC, NASA Glenn Research Center
- Joshua Stuckner
  - Formerly NASA Glenn Research Center
  - Machine Learning Scientist at Cleveland Clinic
- Mark Turner
  - NASA Glenn Research Center
  - University of Cincinnati's Gas Turbine Simulation Laboratory
- NASA Advanced Air Transport Technology Project (AATT)

Topics

1. Surrogate Model Accuracy Study
   - This work is presented in greater detail in a 2025 AIAA Aviation paper
   - ***Structural Optimization of Turbomachinery Rotors: A Study of Machine Learning Surrogate Models***

2. Bayesian Optimization
   - This work will be presented in greater detail in a 2026 AIAA SciTech paper
   - ***Bayesian Optimization Applied to Turbomachinery Structural Optimization***

Surrogate Model Accuracy Study



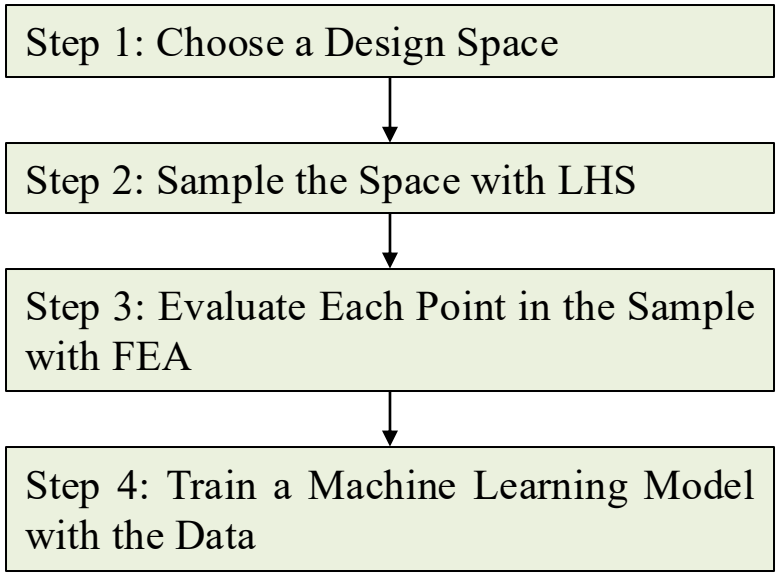
Glenn Research Center

Surrogate Model Accuracy Study

- Many researchers and designers have utilized machine learning surrogate models for design optimization.
- The focus of this work is the study of the accuracy of surrogate models.
  - Comparison of Neural Networks (multilayer perceptron (MLP) type) and Random Forest Models
  - Evaluation of the effect of the Neural Network architecture on accuracy
    - 1, 3 or 5 hidden layers
  - Evaluation of the effect of sample size
    - Ranging from 500 to 5000 samples
  - Evaluation of the effect of the design space dimensionality
    - 4, 6 or 8 design variables (DVs)
  - Evaluation of stochastic variability with 100 repeated model generations
    - Each model is generated with a unique ‘random seed’ parameter
- The surrogate model accuracy is first evaluated using a test data set that was not in any part of the model training.
- Finally, the surrogate models are utilized for design optimization.
  - The surrogate model’s prediction is compared to FEA simulation.

Sampling and Training Methodology

- Training sets are generated using Latin Hypercube Sampling (LHS) with sizes of 500, 1000, 2000 and 5000.
  - A corresponding LHS validation set is generated for each training set.
- Blade geometry is generated using T-Blade3.
  - TBlade3 is a parametric geometry generation tool created at the University of Cincinnati's Gas Turbine Simulation Laboratory.
- For each design space, a test set is generated that contains 2000 samples.
  - The test set is never used in any portion of model training. It is used exclusively for accuracy evaluation.
- The MLP models employ early stopping to avoid overfitting of data.
- The models are trained using margin of safety (MS), which is output by the FEA simulation.



$$\mathrm{MS} = \frac{\text{Yield Strength}}{1.1 \cdot \text{Max Stress}} - 1$$

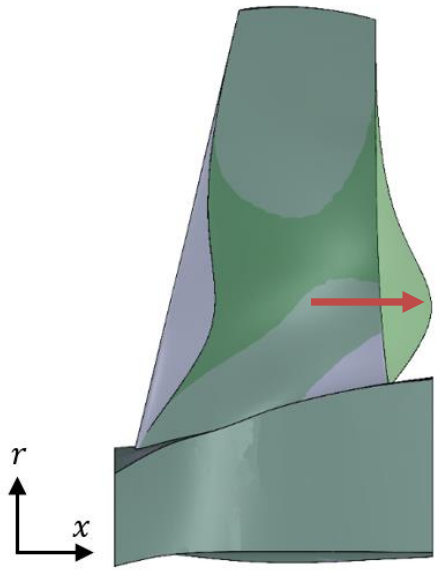
The Design Spaces

- Three design spaces are studied. The design variables in each design space are lean and sweep control points.
- The 4 DV space contains control points at 50% and 100% span locations.
  - The 6 DV space: adds 25% span
  - The 8 DV space: adds 75% span
- Each design variable has the same range: (-0.17,0.03)
  - This range was chosen based on prior optimization results from a previous paper. It is expected that the best design lies within this range.
- The control point at 0% span is always held at 0.

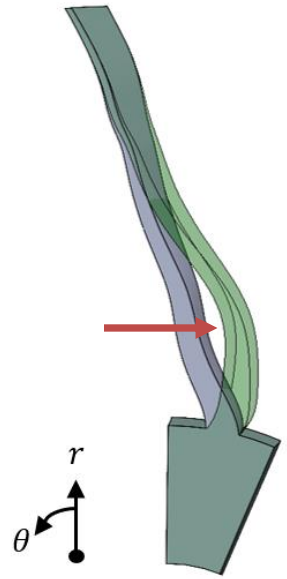
Design Space 1	
Variable	Range
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

Design Space 2	
Variable	Range
$\Delta m'_{25}$	(-0.17,0.03)
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{25}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

Design Space 3	
Variable	Range
$\Delta m'_{25}$	(-0.17,0.03)
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{75}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{25}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{75}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)



Sweep ($\Delta m'_{25}$)



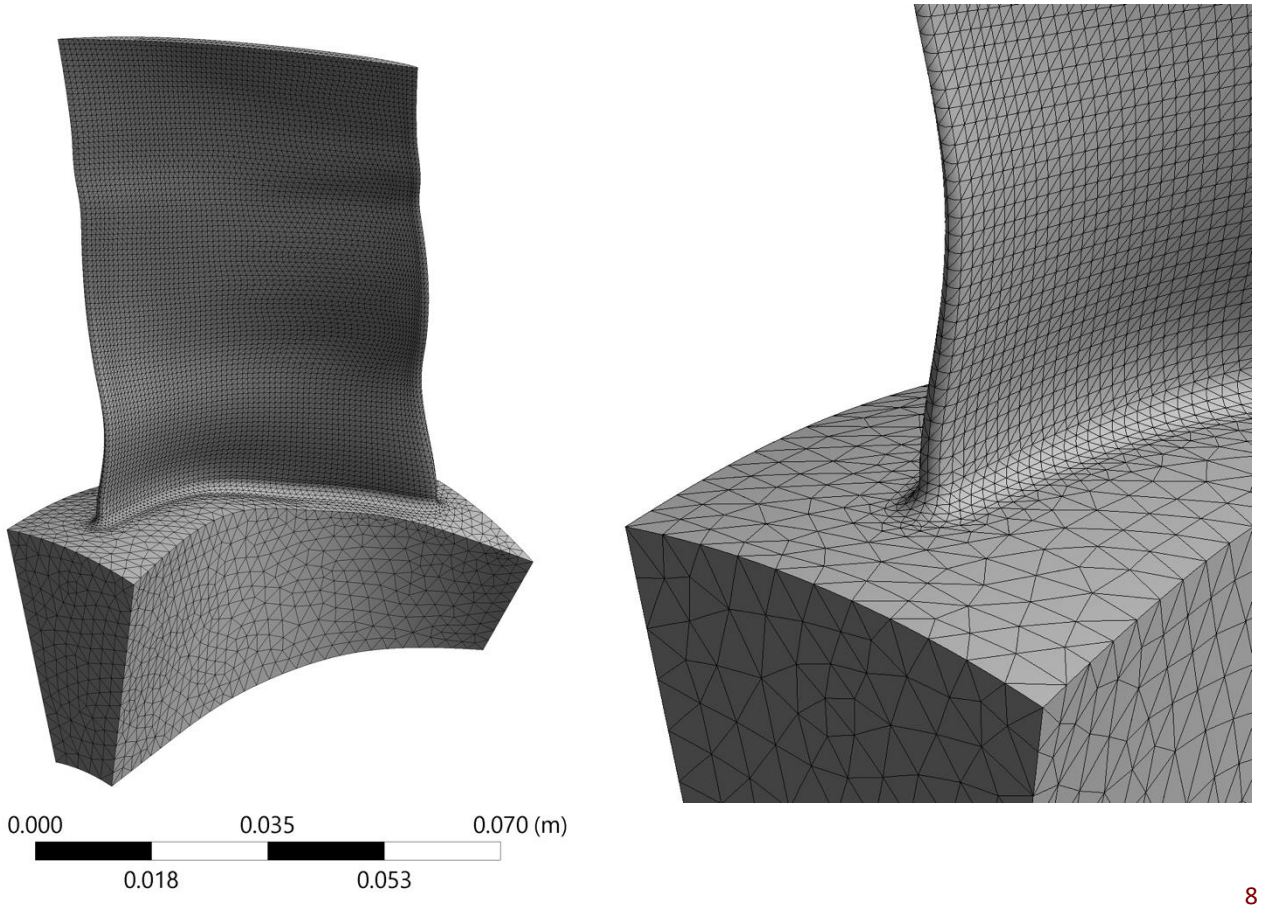
Lean ($\Delta\theta_{25}$)

Baseline Model

- The baseline model is the tail cone thruster 7-28-8.
  - Designed by University of Cincinnati students in collaboration with NASA GRC
  - Openly available on T-Blade3 GitHub page
- All lean and sweep control points are set as 0 for the baseline design.
- Titanium is used as the material.
- The rotor has a diameter of approximately 12 inches.
- The rotational velocity is 21581 RPM.
- A $^{1}/_{16}$ inch fillet is added to the model in Ansys DesignModeler.
- The mesh parameters resulted in a mesh that had approximately 100K elements.
- 1000 designs from the LHS sample can be simulated in a single day.

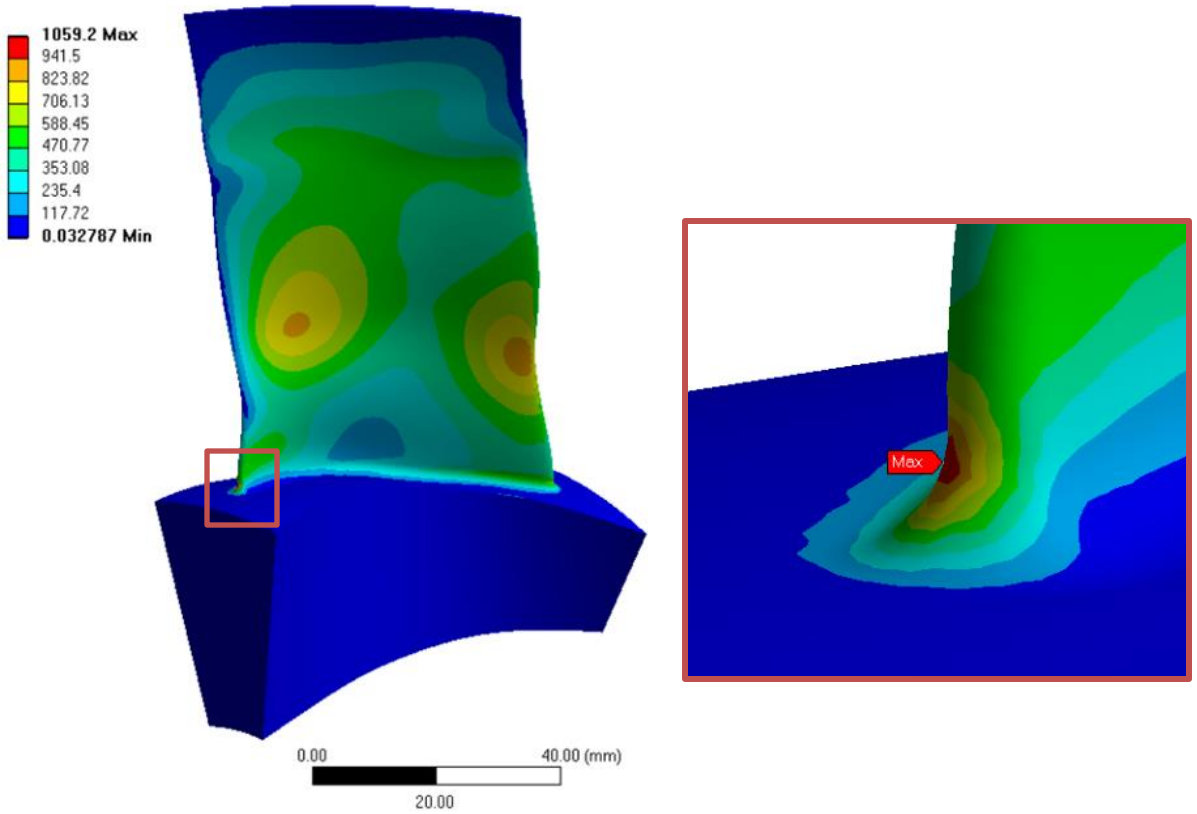
Material Properties of Ti-6Al-4V (Grade 5)

Density [kg/m^3]	4430
Young's Modulus [Pa]	1.14E+11
Poisson's Ratio	0.33
Tensile Yield Strength [Pa]	1.10E+09



Baseline Model Results

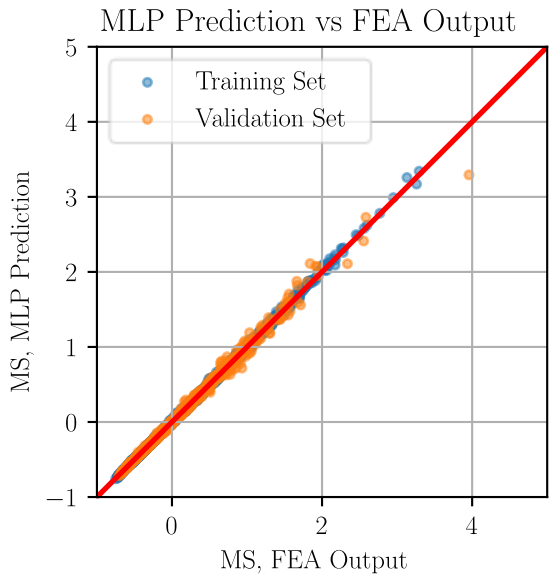
- The analysis of the baseline design resulted in a margin of safety of -0.0545.
- The max stress occurs in the fillet region.

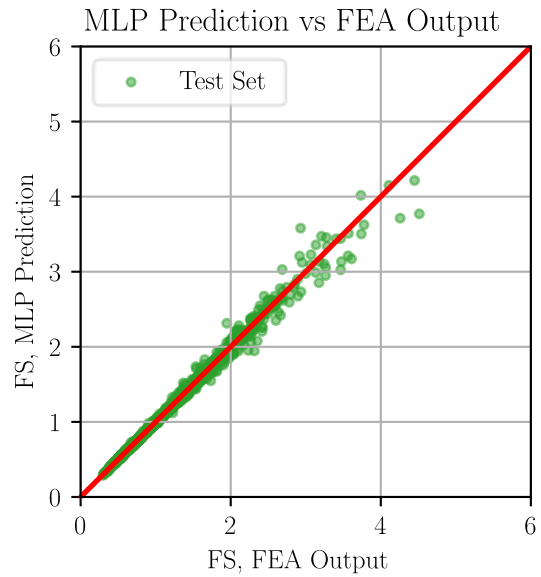


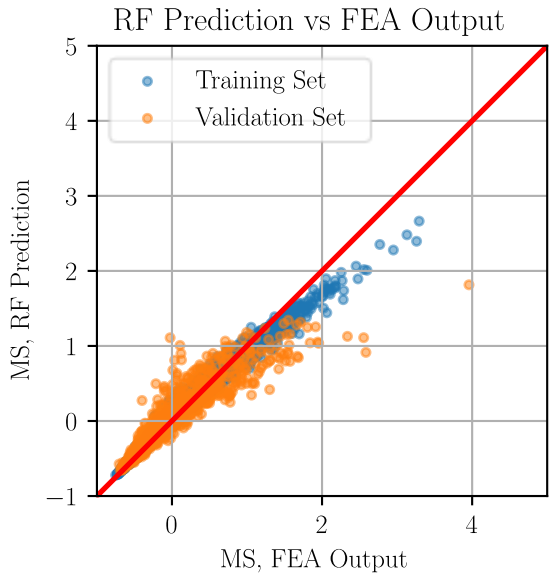
Accuracy Evaluation Metric

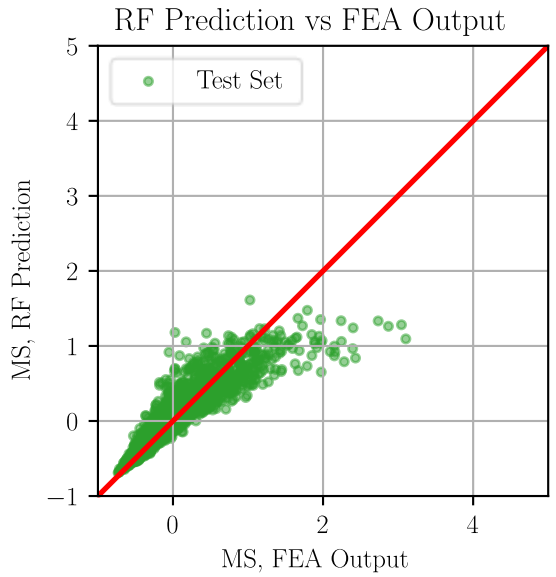
- These plots were generated using models trained with 5000 samples from the 6 DV design space.
- Models are initially evaluated using the coefficient of determination (R^2) metric that is often used to assess the accuracy of trendlines.
- The plots to the right show MS values predicted by the surrogate model compared to MS values from the FEA.
  - **x-axis:** FEA result
  - **y-axis:** prediction from the surrogate model
- Ideally, the FEA result and the surrogate model prediction would correspond one to one and fall directly on the red line.

	MLP	RF
Training Set R^2	0.9993	0.9751
Validation Set R^2	0.9944	0.8175
Test Set R^2	0.9919	0.8076



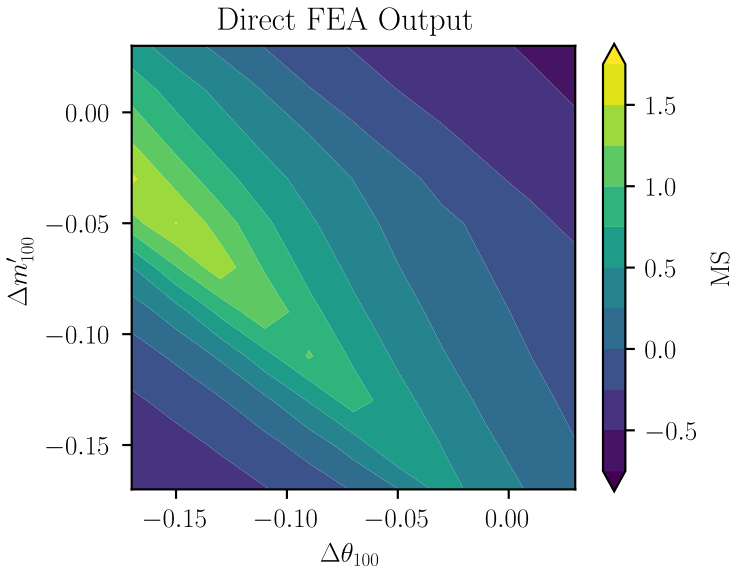


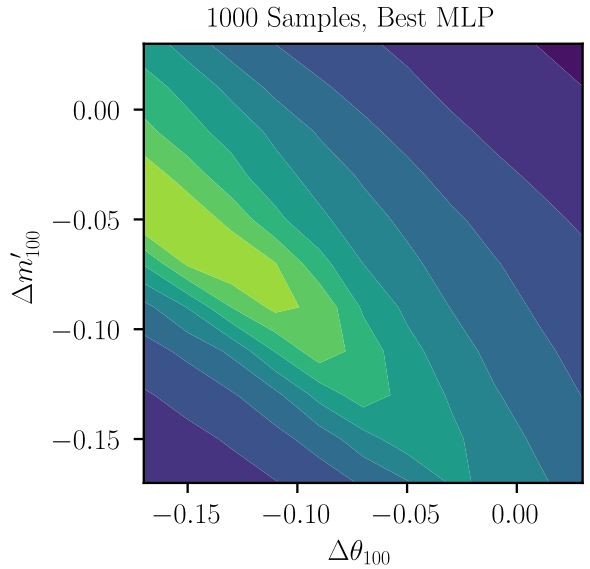


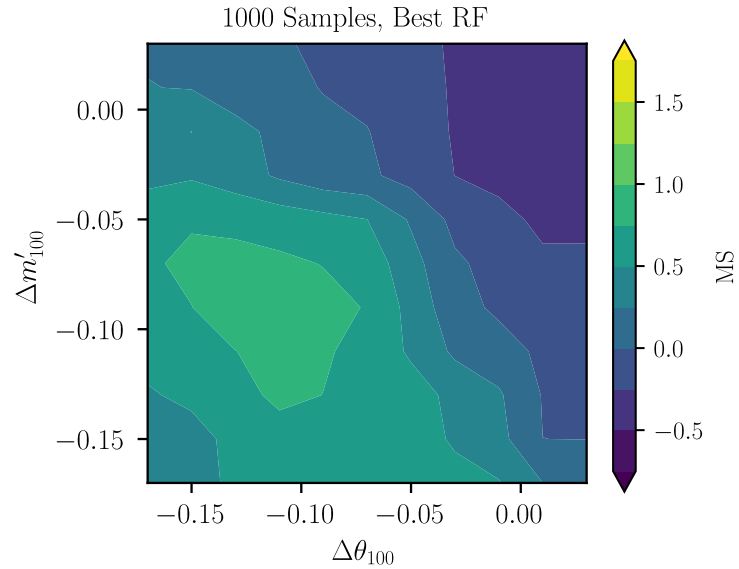


6 DV Surrogate Model Visualization

- The visualization to the right reinforces the higher accuracy of the MLP models compared to the RF models.
  - Generated using surrogate models trained with 1000 samples
  - Consistent with the R^2 evaluation
- The plots are a slice of the six-dimensional design space.
  - $\Delta m'_{100}$ and $\Delta\theta_{100}$
    - Varied across their whole range
  - All other variables
    - Fixed at the center of the design space

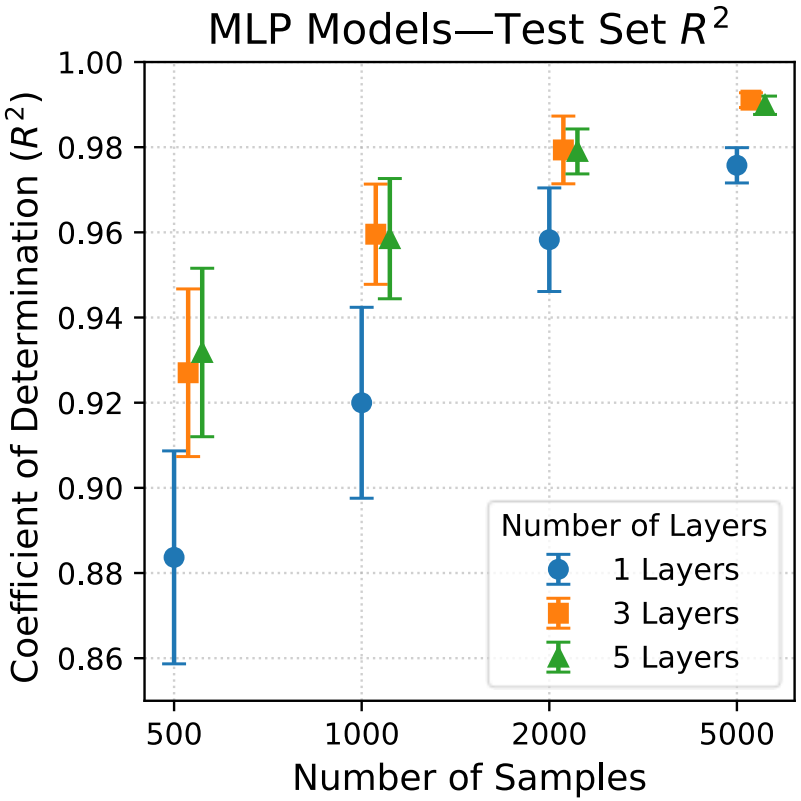






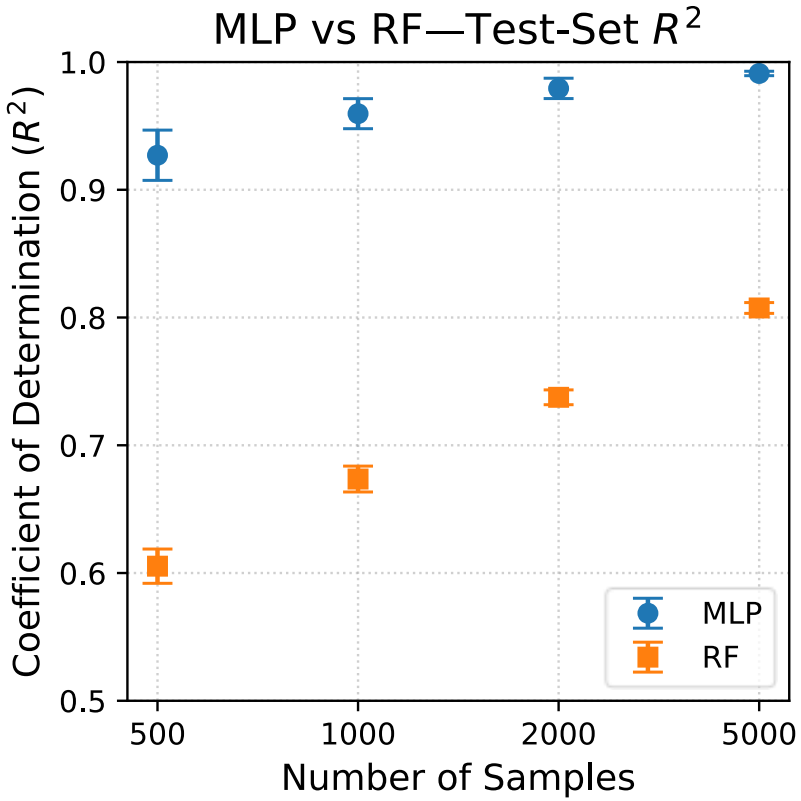
6 DV Design Space - MLP Evaluation

- Each type of machine learning model is trained 100 times using 100 different random seeds.
  - Neural network model training has inherent randomness built into it through weight initialization and data shuffling.
- The models are evaluated using the test dataset.
  - Each marker is the mean, with error bars showing a range of 2 standard deviations (95% of cases).
- The performance of the MLP models showed a wide variance, even when model type and training samples were the same.
  - This indicates that it may be beneficial to train several MLP surrogate models and select the best one based on the validation fraction accuracy.
- The performance of the three-layer MLP models are significantly better than the one-layer model.
  - Five-layer models are similar to three-layer models.



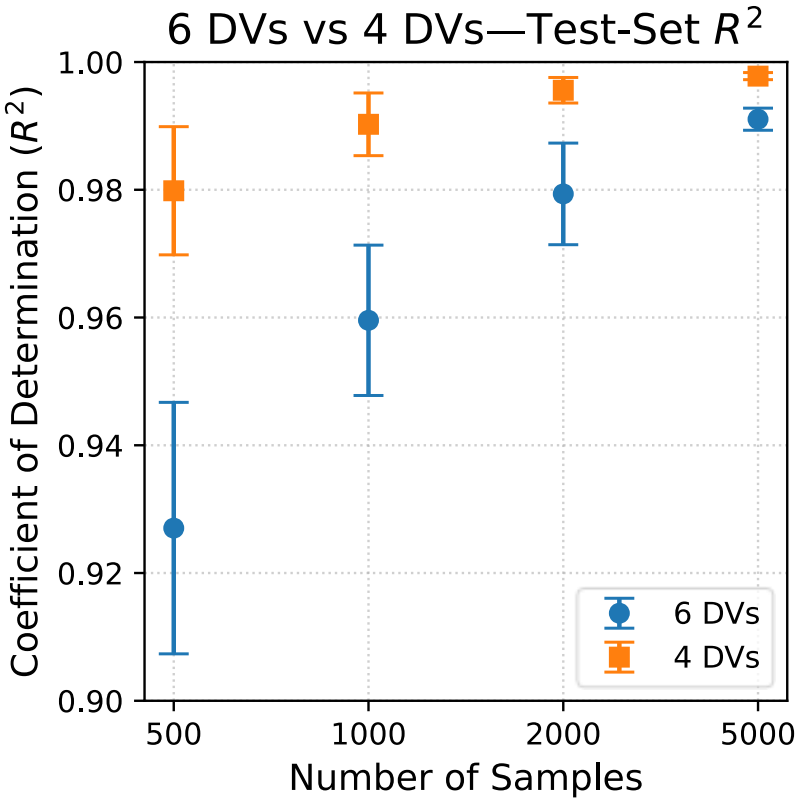
6 DV Design Space - Random Forest

- In terms of R^2 value, the RF models perform poorly compared to the neural network models.
- The best performing RF model trained with 5000 samples is less accurate than the MLP models trained with only 500 samples
  - This is consistent across all three design spaces.
- In the interest of time, RF model accuracy evaluation using R^2 will not be shown for the rest of this presentation
  - Similar evaluations for 4 DVs and 8 DVs can be found in the Aviation 2025 paper that details this study.



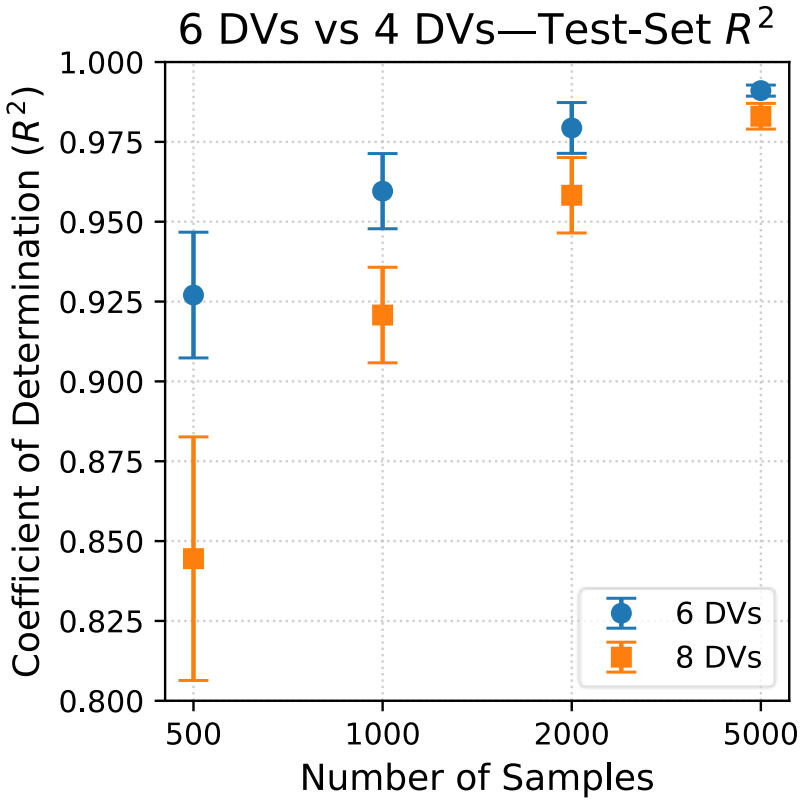
4 DV MLP Evaluation

- All accuracy evaluation figures from this point forward assess only the 3-layer models.
- As expected, the surrogate models for the 4 DV space attain higher accuracy compared to the 6 DV space.
  - This was most significant for the models trained with 500 samples.
  - The effect was less pronounced for the models trained with 5000 samples.
- The variance in model accuracy is also less for the 4 DV space compared to the 6 DV space.
  - Recall that these are identical models other than the difference is the random seed.
- The accuracy again improves when moving from one to three layers, but no significant improvement is seen when moving from three to five layers.



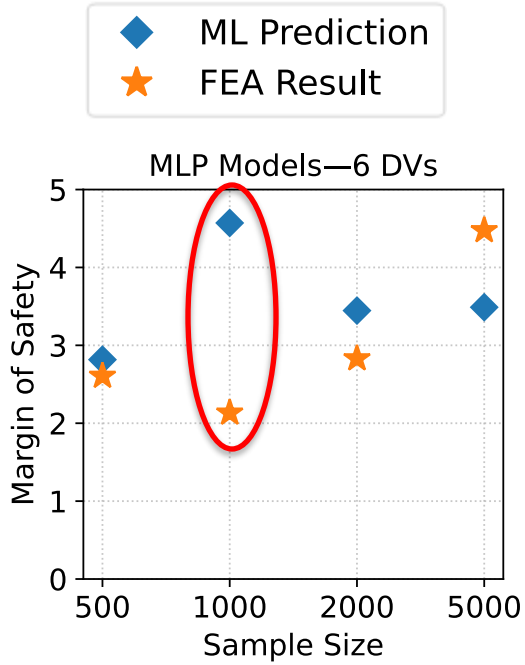
8 DV MLP Evaluation

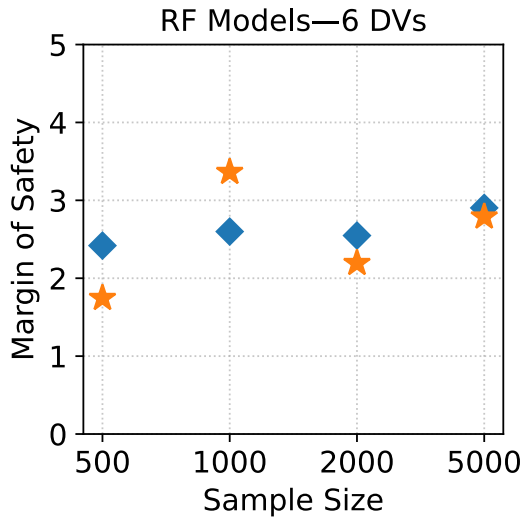
- The drop-off in model accuracy begins to become more pronounced for the 8 DV space.
  - This is most evident in the models trained with less samples.
- There is more variance in the model accuracy for the higher dimensional space.
- Like the other two design spaces, the accuracy improves when moving from one to three layers, but no significant improvement is seen when moving from three to five layers.



Design Optimization, 6 DVs

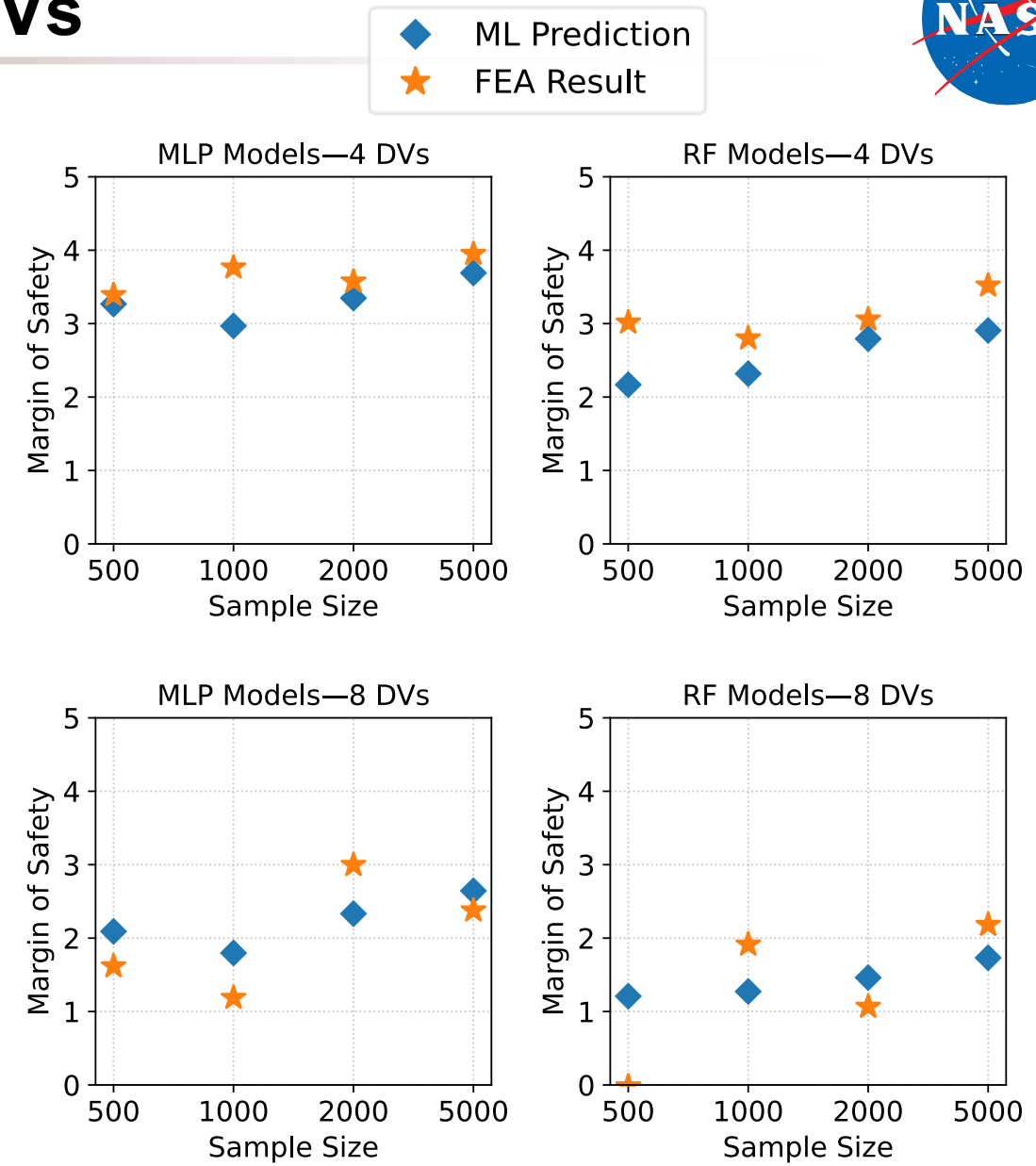
- The best performing surrogate model in terms of validation set accuracy is chosen for this evaluation.
- In this evaluation, the aim is to determine whether the model can predict the optimal location within the design space.
- The optimization is conducted using a method called Differential Evolution.
  - Available in the Python package SciPy
  - Default parameters are used.
- The optimization is run 100 times.
  - Differential Evolution relies on random processes during its search.
  - Most of the optimization runs resulted in the same design while some were stuck in a local minimum.
- The optimal point predicted by the surrogate model is then evaluated using FEA.
- MLP models generally outperformed the RF models, however, one MLP model suffered from extrapolation issues.





Design Optimization, 4 DVs and 8 DVs

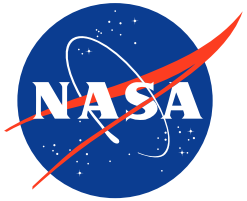
- The surrogate models trained in the space with 4 DVs produced the most consistent results.
- The 4 DV designs underperform the best result from 6 DVs.
  - This is likely a limitation in the types of geometries that can be produced using 4 DVs.
- For the 8 DV space, the surrogate models appear to suffer from the ‘curse of dimensionality.’
  - The larger the number of inputs to the surrogate model, the more model evaluations are needed to construct a surrogate model that is accurate enough.
  - In this case, it seems that even 5000 training points is not enough for the model to accurately represent the space.
  - The best MLP models for the 8 DV space underperformed compared to the worst performing MLP model from the 4 DV space.
- The RF models generally underperformed compared to the MLP models.



Surrogate Modelling Conclusions

- MLP models consistently outperformed RF models, achieving higher accuracy (R^2) on both test and validation sets.
- Generating multiple MLP models with identical architecture is advantageous.
  - The best model can then be selected using the validation set.
- Despite superior accuracy, MLP models experienced some extrapolation issues during optimization, occasionally predicting unrealistic optima at the edges of the design space.
- Accuracy significantly declined with increased dimensionality (curse of dimensionality).
  - This resulted in poorer performance in design optimization.
- As a next step in this research, focus will be shifted to Bayesian Optimization methods, which avoid spending excessive resources in regions of poor performance.
- A comparison of the optimized design and the baseline design will be shown at the end of the next section.

Bayesian Optimization



Glenn Research Center

Bayesian Optimization

- Bayesian optimization offers an intelligent means of selecting points to build the surrogate model.
- **Focuses sampling where it matters**
  - prioritizes regions likely to improve results.
- **Works in high-dimensional spaces**
  - Effective even when design spaces are complex.
  - Works well with up to 15 design variables.
- **Adaptive and data-driven**
  - Surrogate model updates with each new trial.
- **Applicable to various types of design variables**
  - Continuous Variables (range of floating-point numbers)
  - Integer Variables (range of integers)
  - Categorical Variables (labels without inherent order)
    - For example: red, green, blue
    - Material selection is a good example of a categorical variable that could be used in turbomachinery design.

Bayesian Optimization Basics

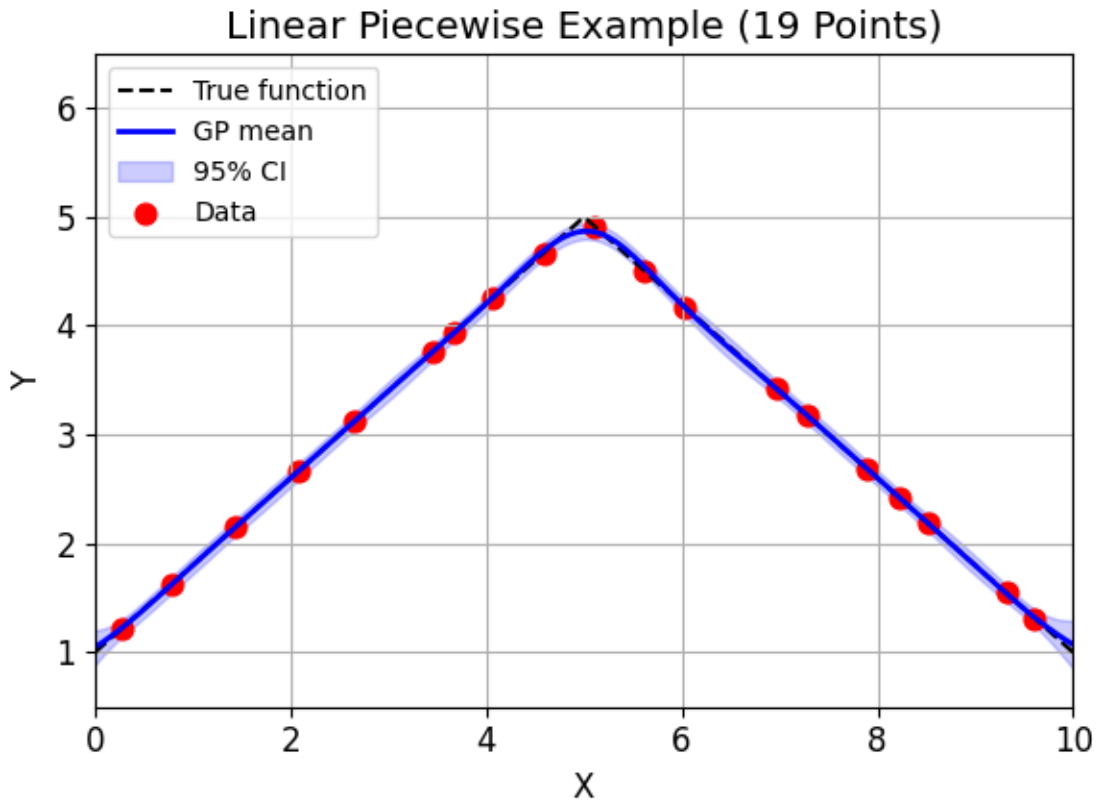
- **Surrogate Model**
  - Most often a Gaussian Process regression model, a type of probabilistic surrogate model
  - Returns an expected mean and uncertainty throughout the domain
  - Other surrogates can also be utilized such as a Monte Carlo or Bayesian Neural Network.
    - The application of alternative surrogates is being considered for future research.

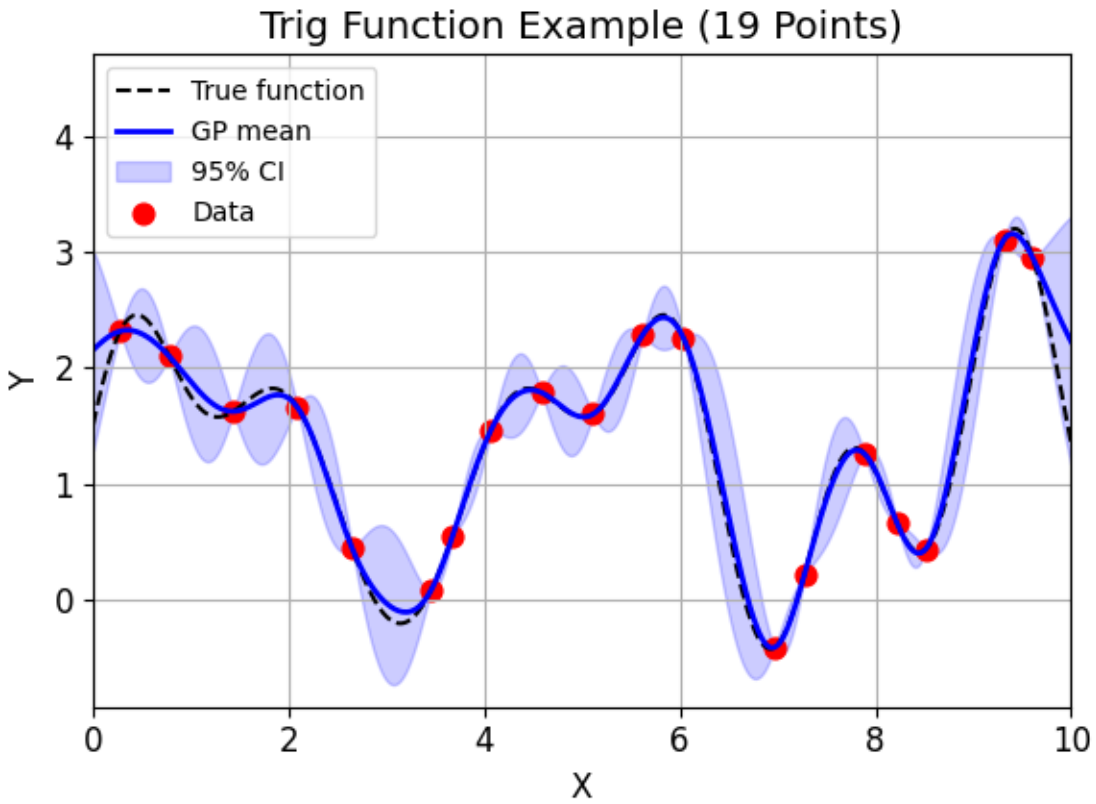
- **Mean Value**
  - The best estimate of the functions value based on existing samples

- **Standard Deviation**
  - An estimate of uncertainty based on existing samples

Examples Problems

- Two example problems will be used to demonstrate Bayesian optimization.
  - These are first demonstrated on this slide using points selected a priori using LHS.
  - The 95% Confidence Interval (CI) is determined by $\mu \pm 1.96\sigma$.
    - μ: mean
    - σ: Standard Deviation

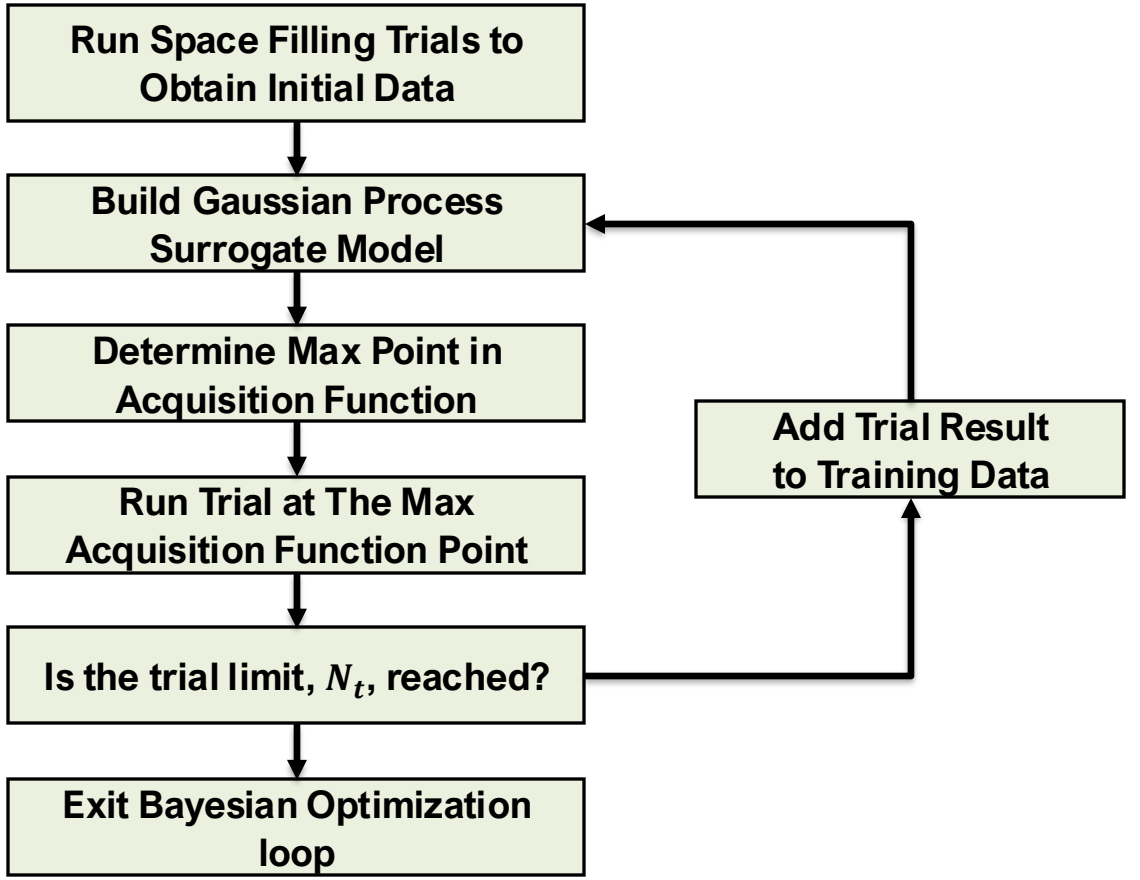




Acquisition Functions

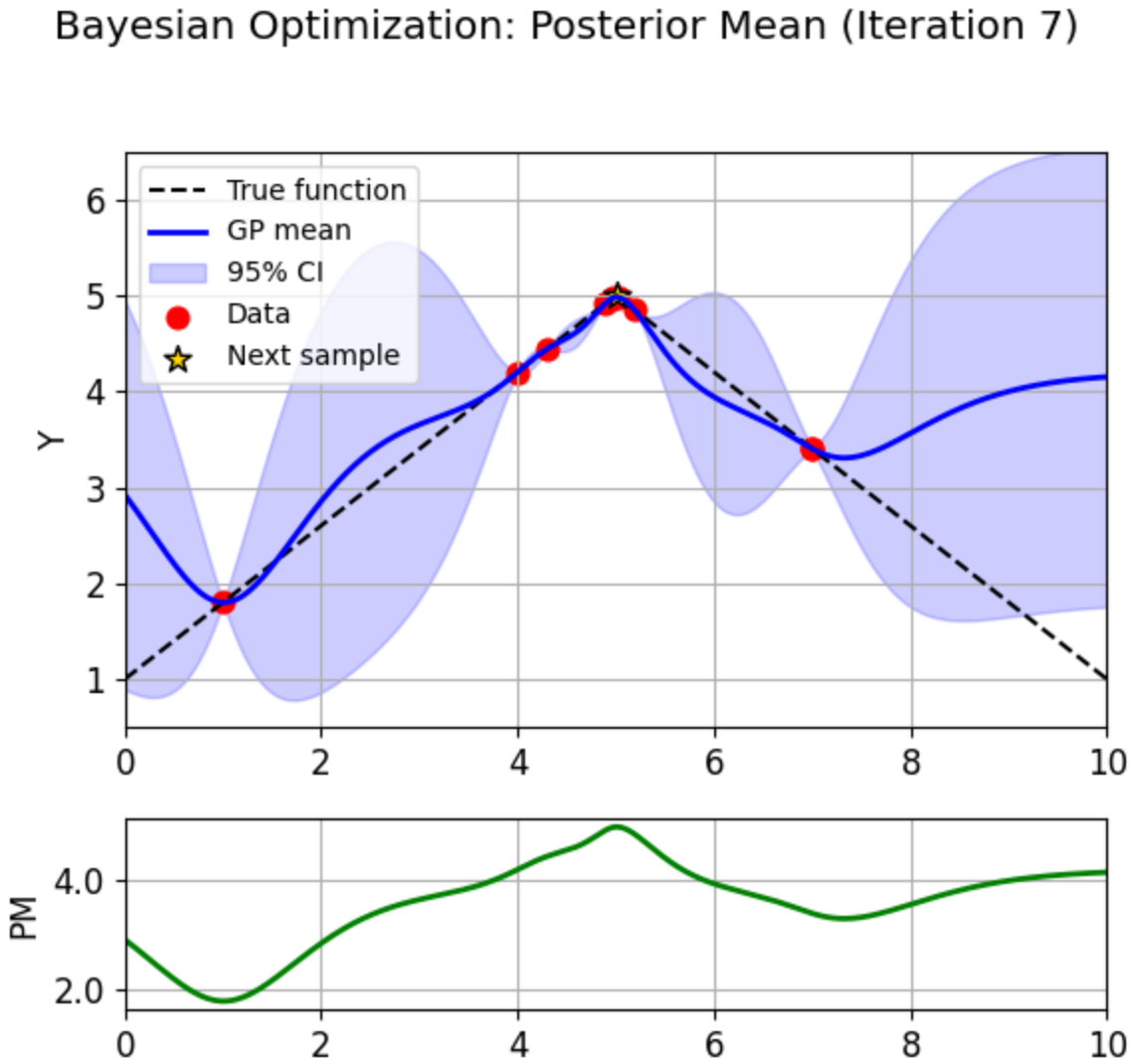
- The surrogate model requires data to provide mean predictions and uncertainty quantification.
  - A typical optimization begins with space filling trials to generate initial data before use of the acquisition function.

- Acquisition Function
  - The acquisition function typically uses the mean and standard deviation metrics to determine the next point to sample.
  - It can be adjusted to favor areas of high mean or areas of high uncertainty.
    - Exploration vs. Exploitation
- There are many types of acquisition functions. Three are described in this seminar.
- The outputs from the Gaussian Process Regressor are:
  - The predicted mean: $\mu(x)$
  - The uncertainty in terms of standard deviation: $\sigma(x)$



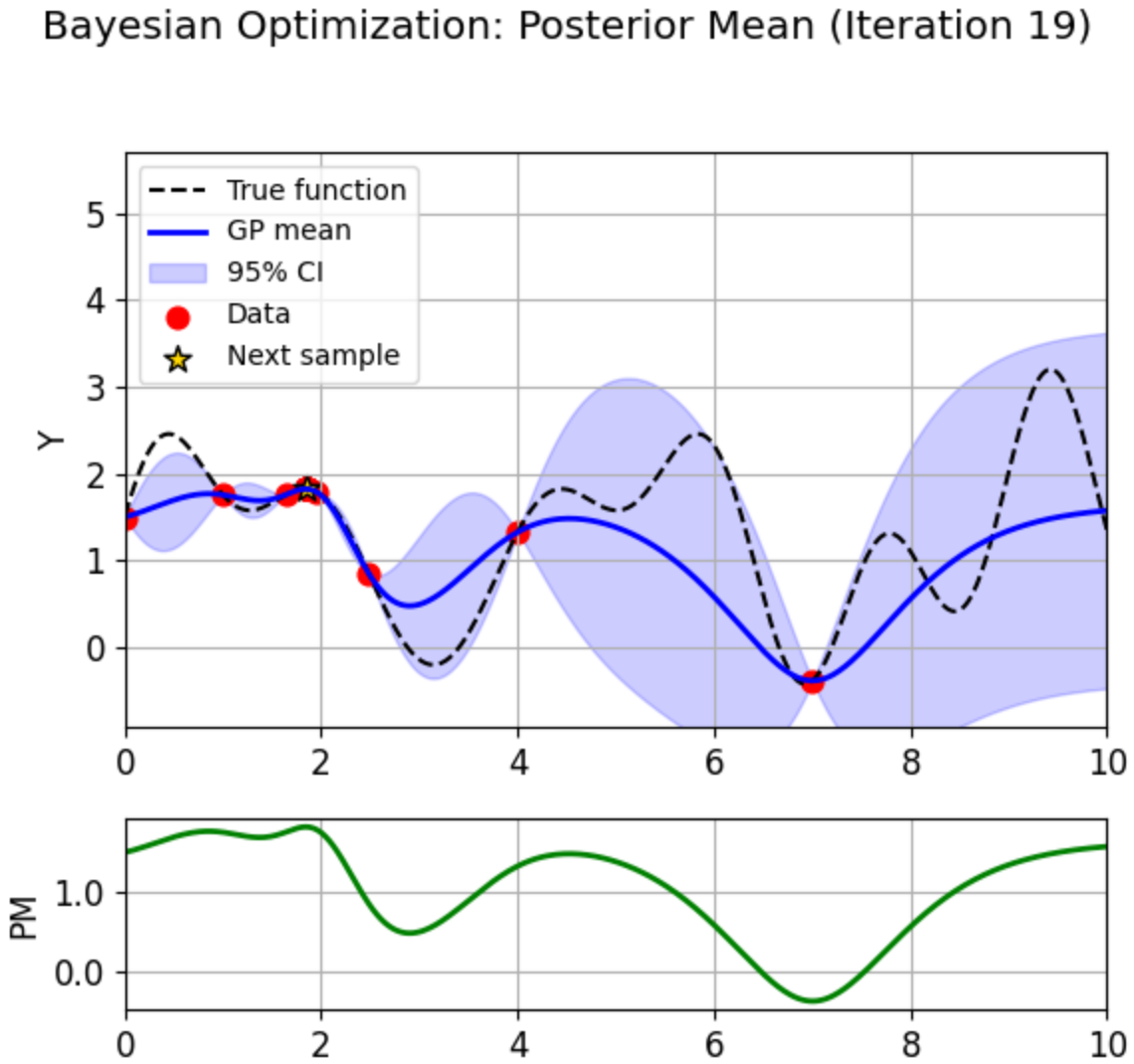
Posterior Mean: Linear Piecewise Case

- Posterior Mean
  - $\mathrm{PM}(x) = \mu(x)$
  - Simplest acquisition function
  - Does not consider uncertainty
- For the simple linear piecewise example, the PM method rapidly finds a value that is very near the maximum.
- After the seventh iteration, the PM method converges on a value and continuously provides the same location as the next trial point.
- Although there are regions of high uncertainty, they are not explored. For this case, that is not an issue.



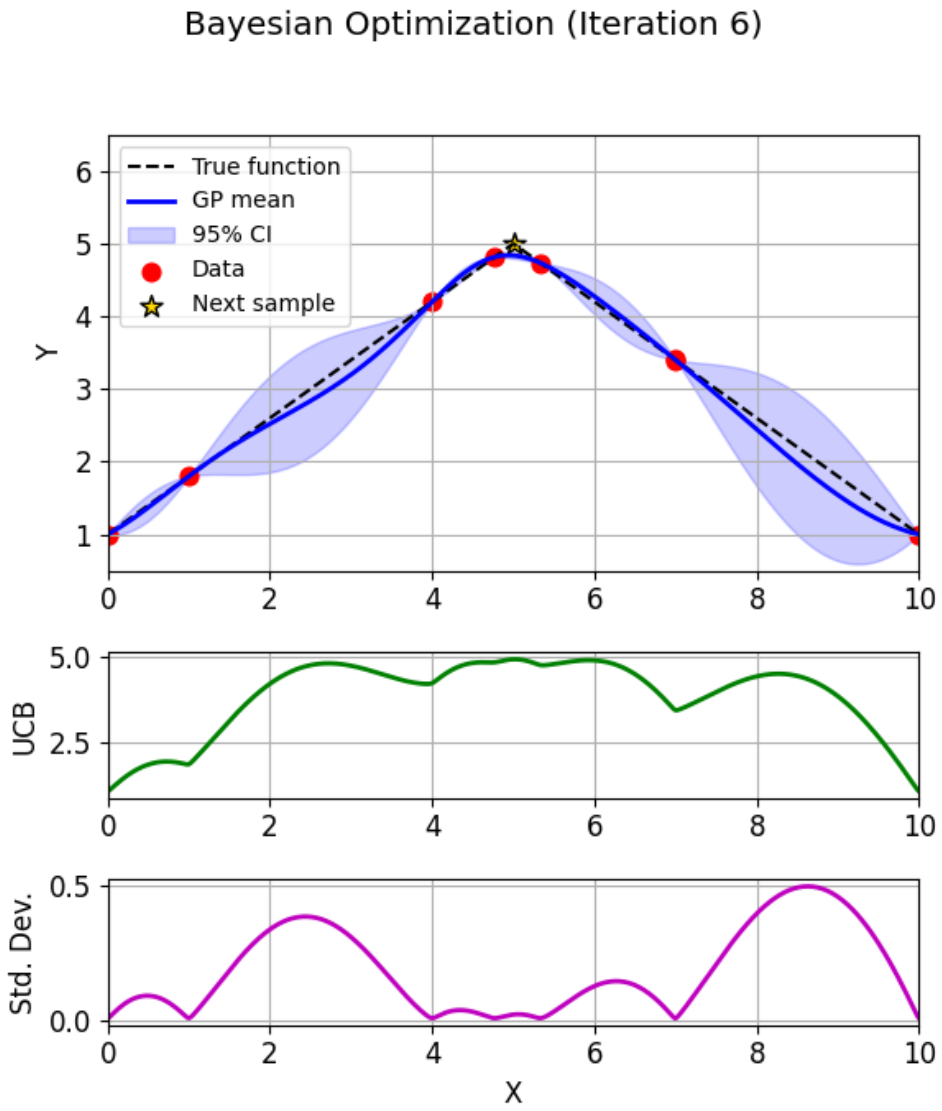
Posterior Mean: Trig Function Case

- The PM acquisition function performed well on the linear piecewise case.
- However, it has a very high likelihood of becoming stuck on local maxima since it does not consider uncertainty.
- In this case, the method converged on a point after 4 iterations and repeatedly recommended the same point afterwards.
  - As expected, PM provided very little exploration as it does not consider uncertainty.
  - This point was not the maximum in the domain!



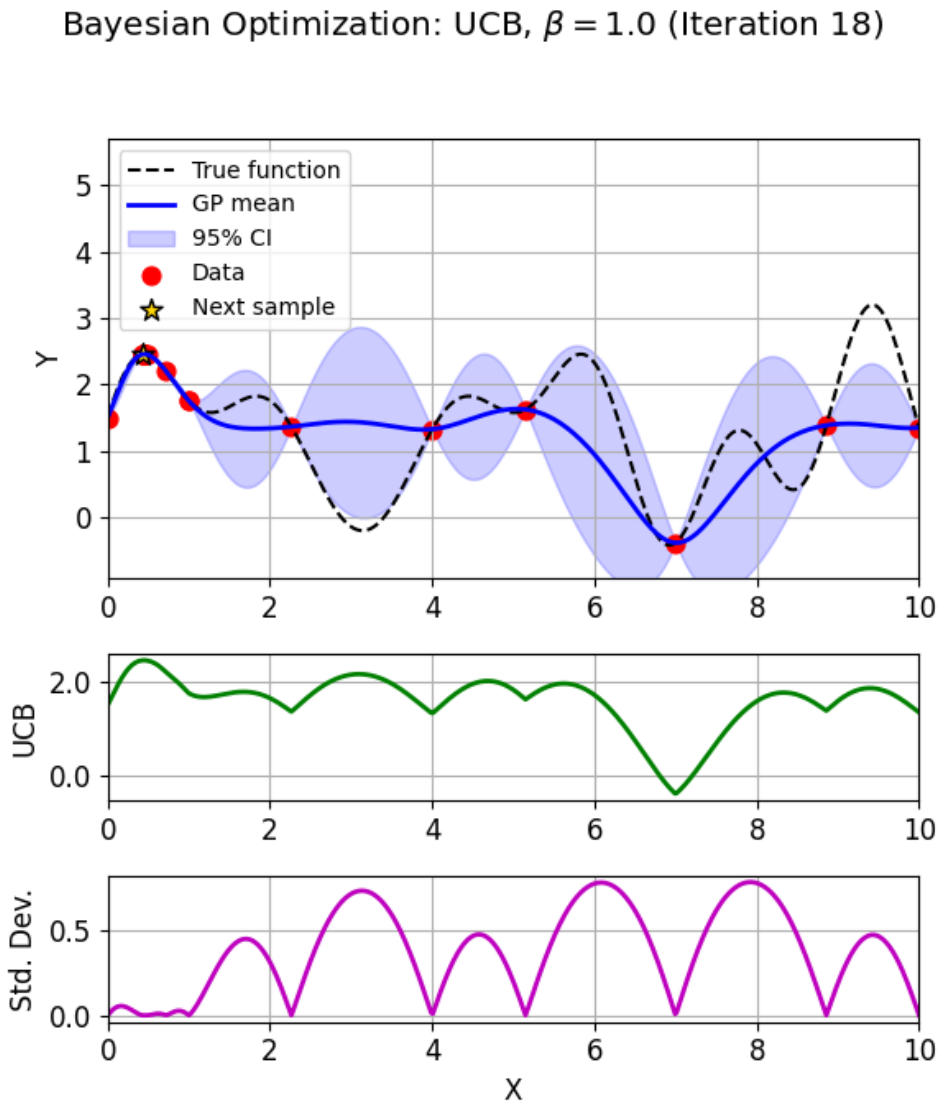
Acquisition Function: Upper Confidence Bound

- Upper Confidence Bound
  - $UCB(x) = \mu(x) + \beta\sigma(x)$
  - Incorporates uncertainty in a straightforward way
  - β is a user specifiable parameter that weights the mean and the uncertainty.
    - Weighting uncertainty higher (larger β) causes the method to be more exploratory.
  - For this example, $\beta = 5.0$
    - Exploration focused
    - For this case, a β value of 1.0 performs better, finding the peak after just 2 iterations.



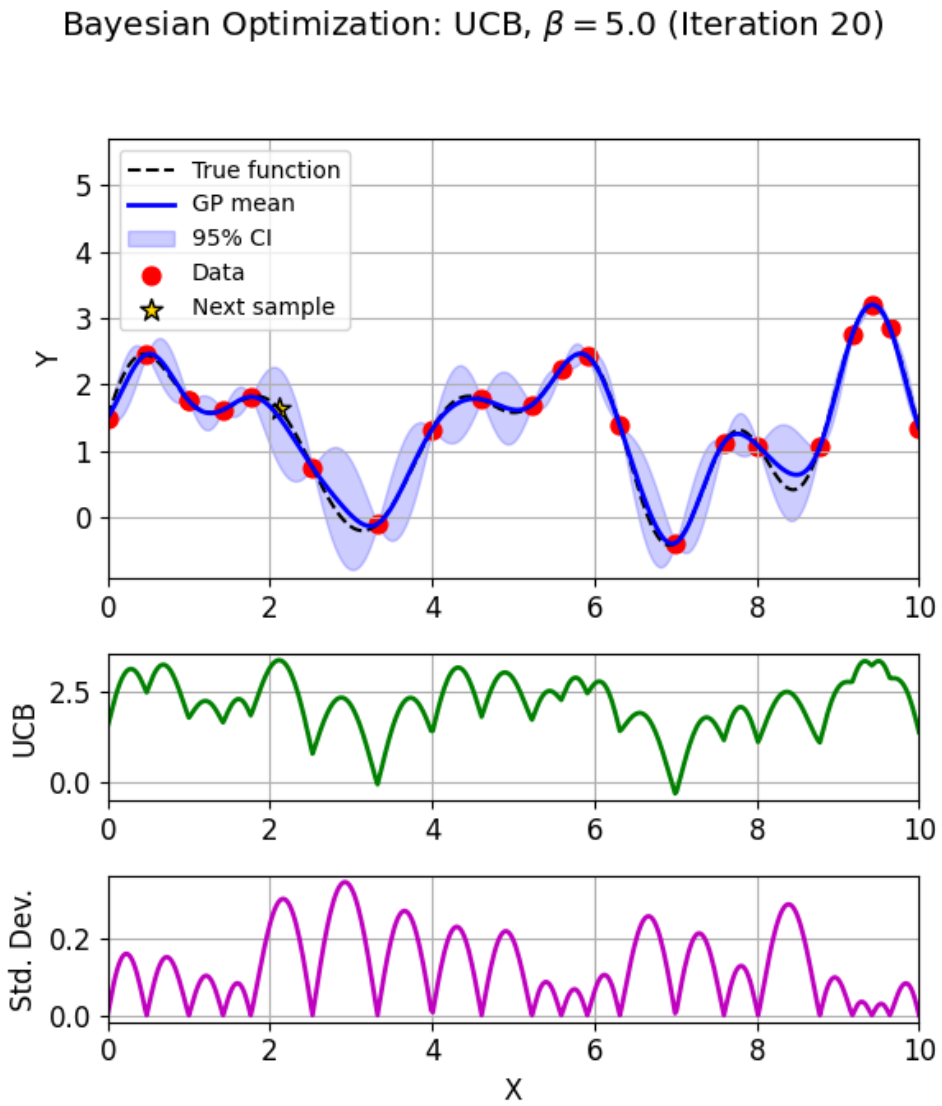
Acquisition Function: Upper Confidence Bound

- The following two examples will showcase the effect of the β parameter on space exploration.
- First, consider the case to the right with $\beta = 1.0$.
- Like the posterior mean case, this case becomes stuck in a local maxima.
  - Repeatedly provides that point as the next trial.
  - After Iteration 17, it is completely stuck.
- This highlights the detriment of not favoring exploration enough.



Acquisition Functions: Upper Confidence Bound

- With $\beta = 5.0$, the UCB function becomes much more exploratory.
- At iteration 12, the best point in the range is proposed as a trial.
- At iteration 20, one can see the highly exploratory nature of the $\beta = 5.0$ setting has thoroughly explored the space.
- Note that there is a detrimental effect of setting β too high in that more resources may be spent in regions of low performance.
- My testing has shown that for the structural optimization problem, $\beta = 5.0$ provides a good balance of exploration.



Acquisition Functions: Expected Improvement

- Expected Improvement
  - $\mathrm{EI}(x) = \begin{cases} (\mu(x) - f^+ - \xi)\Phi(Z) + \sigma(x)\phi(Z), & \sigma(x) > 0 \\ 0, & \sigma(x) = 0 \end{cases}$
  - $Z = \frac{\mu(x) - f^+ - \xi}{\sigma(x)}$

- Upper Confidence Bound
  - $UCB(x) = \mu(x) + \beta\sigma(x)$

f^+: current best design

$\Phi(Z)$: cumulative distribution function (CDF) of the standard normal distribution
- The probability that the prediction exceeds the current best

$\phi(Z)$: probability density function (PDF) of the standard normal
- The weighting from uncertainty

ξ: The exploration parameter
- if $\xi > 0$: the process penalizes exploitation and encourages exploration
- Ax allows only $\xi = 0$

- **Expected Improvement (EI):** Picks points where the *average expected gain* over the best result so far is biggest, factoring in both how good the prediction is and the chance it beats the current best.

- **Upper Confidence Bound (UCB):** Picks points with high predicted mean *or* are very uncertain, depending on the weighting used through the β parameter.

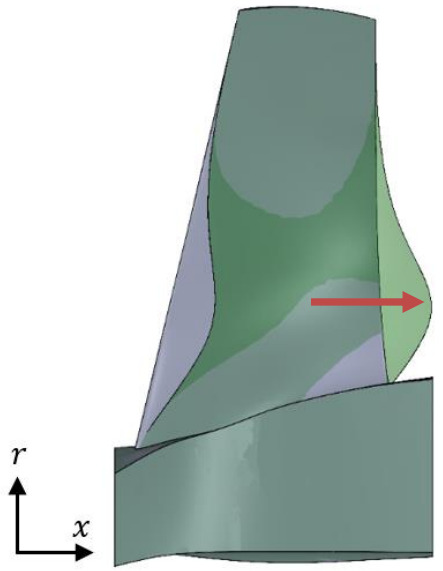
Application to Structural Optimization

- The Bayesian optimization method is now applied to the same optimization problem that was presented in the first section of this talk.
- The method uses 500 trials on each design space.
  - 50 space filling trials (Sobol Sequence)
  - 250 Upper Confidence Bound Trials with β=5.0
    - Highly exploratory
  - 100 Expected Improvement Trials
    - More exploratory, less exploitive
  - 100 Posterior Mean Trials
    - Highly exploitive
- Recall that 500 samples was the smallest sample size used in the Neural Network study presented in the first half of this talk.

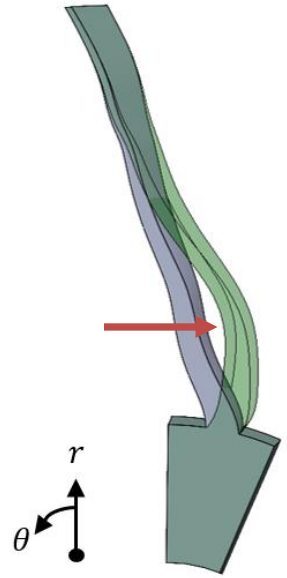
Design Space 1	
Variable	Range
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

Design Space 2	
Variable	Range
$\Delta m'_{25}$	(-0.17,0.03)
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{25}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

Design Space 3	
Variable	Range
$\Delta m'_{25}$	(-0.17,0.03)
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{75}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{25}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{75}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)



Sweep ($\Delta m'_{25}$)

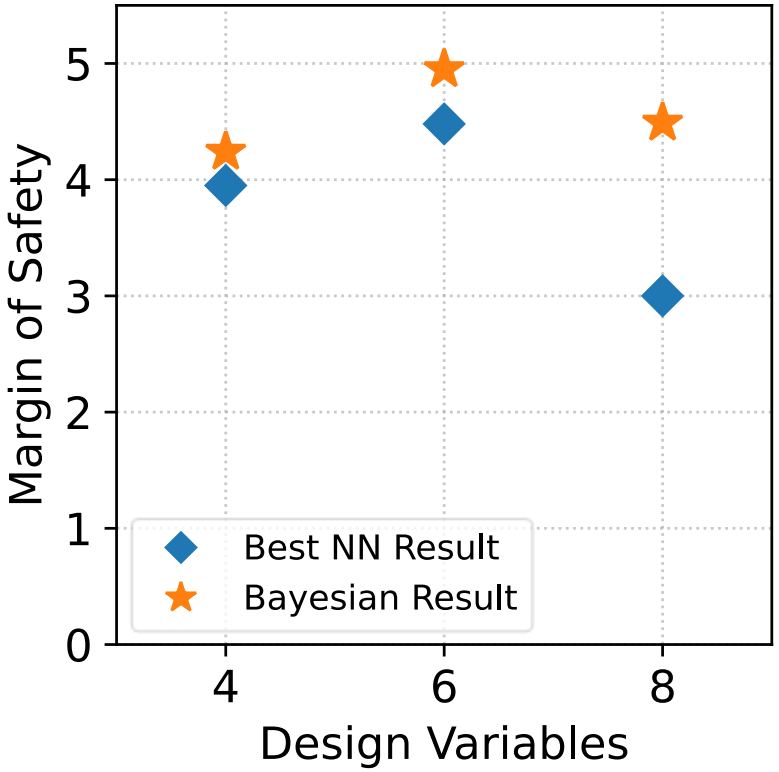


Lean ($\Delta\theta_{25}$)

Bayesian Optimization Results

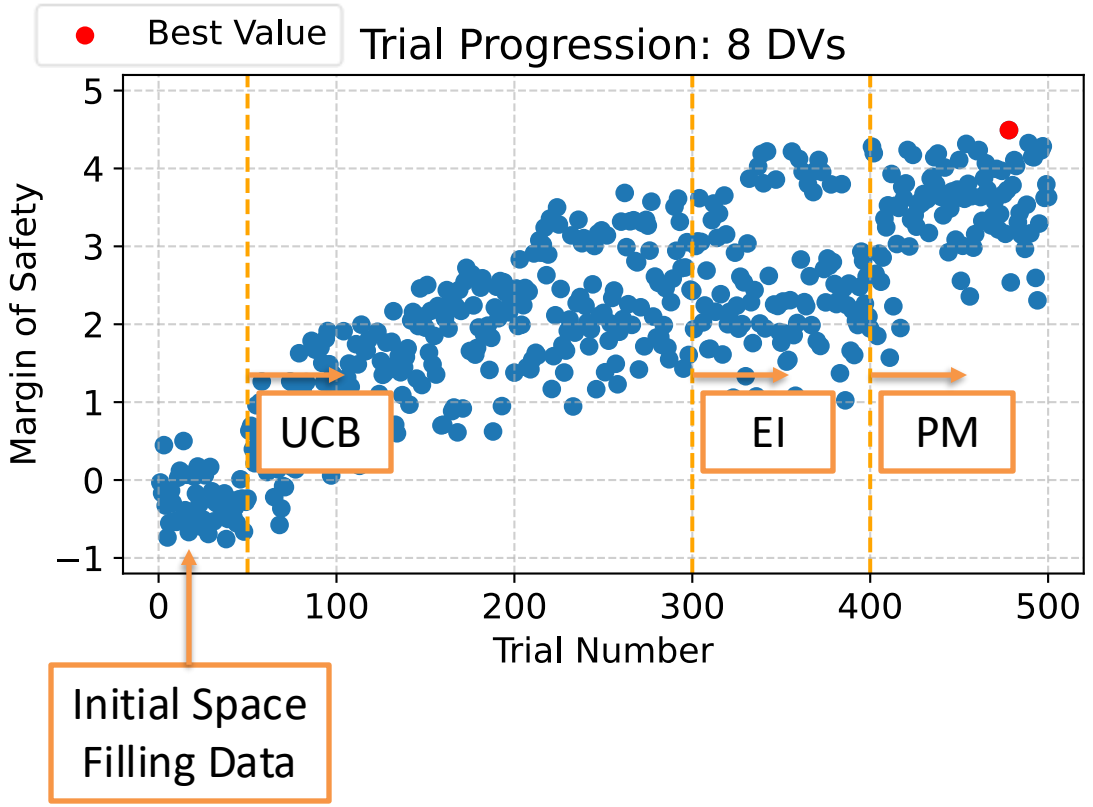
- The best MS results from the neural network surrogate models is shown to the right for each of the three design spaces.
  - Results are the FEA evaluation
  - Significant improvement over the baseline value of -0.0545
- The NN surrogate model results were costly to obtain.
  - 4 DV and 6 DV results obtained from a NN trained with 5000 samples
  - 8 DV results obtained from a NN trained with 2000 samples
- **Bayesian optimization achieved a superior MS result with 500 trials.**

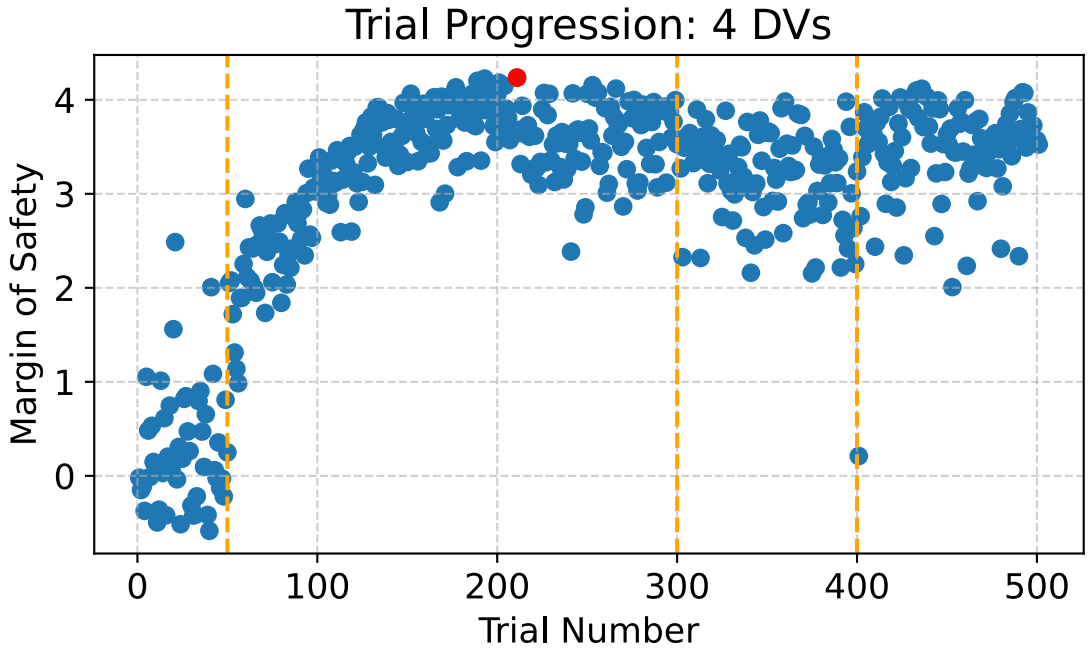
	4 DVs	6 DVs	8 DVs
Best Neural Network Result	3.95	4.48	3.00
Bayesian Optimization Result	4.24	4.95	4.49

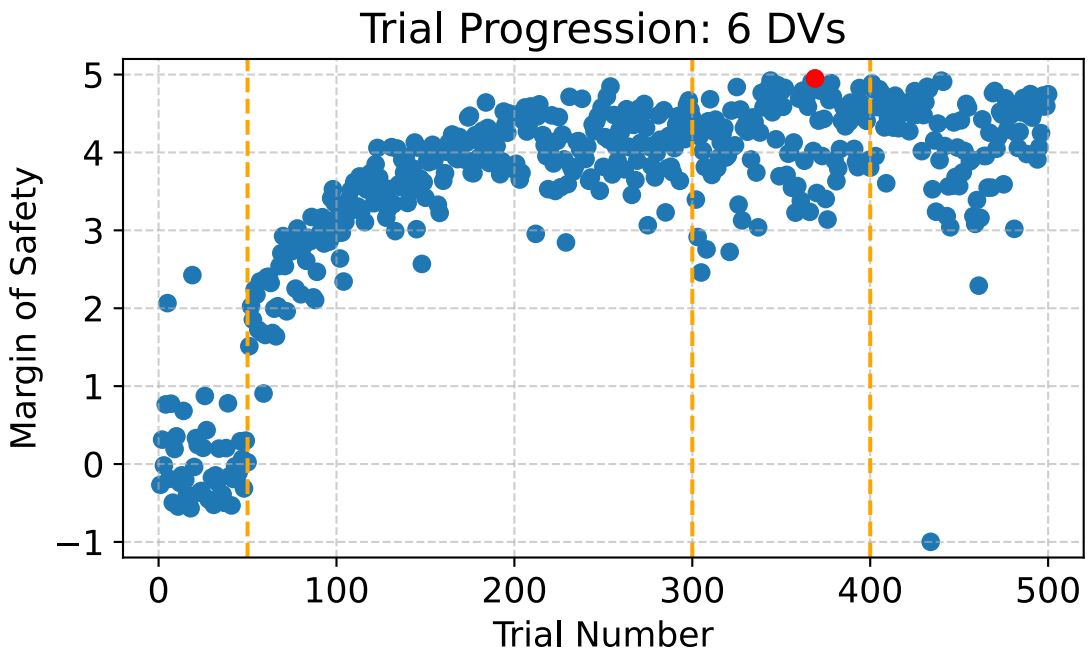


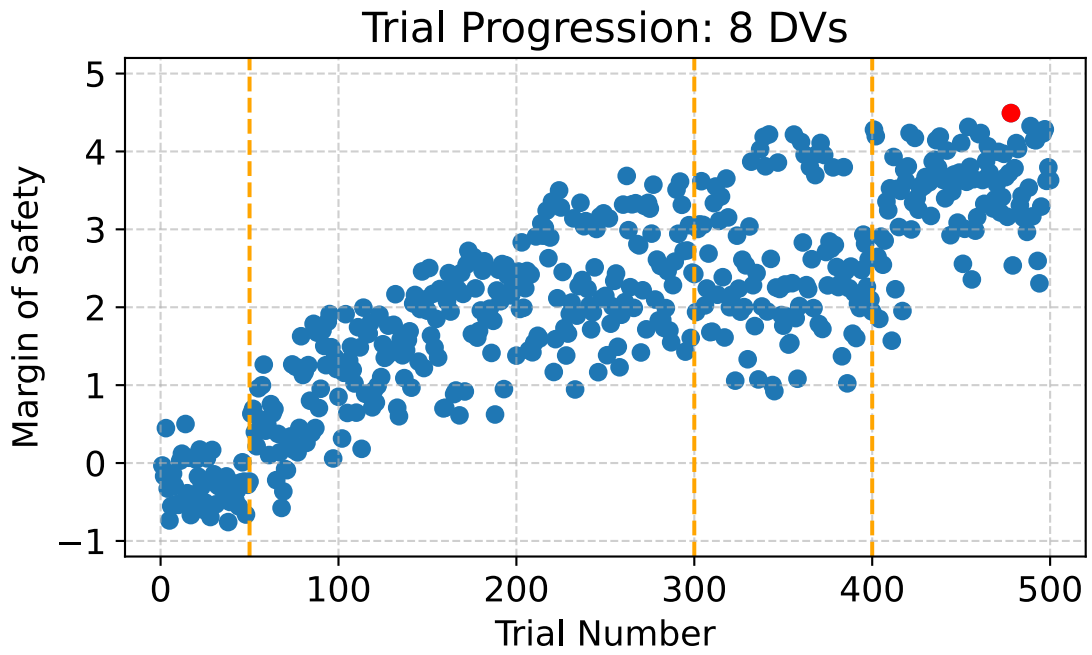
Bayesian Optimization Results

- For the 6 DV space, the best result is found at trial 369 during the EI segment of Bayesian optimization.
- In the 4 DV space, the best result is achieved more quickly at trial 211 during the UCB segment of Bayesian optimization.
- In the 8 DV space, the best result is achieved at trial 478 of Bayesian optimization.
  - Compared to the other results, the 8 DV space appears as if it would benefit from additional trials.



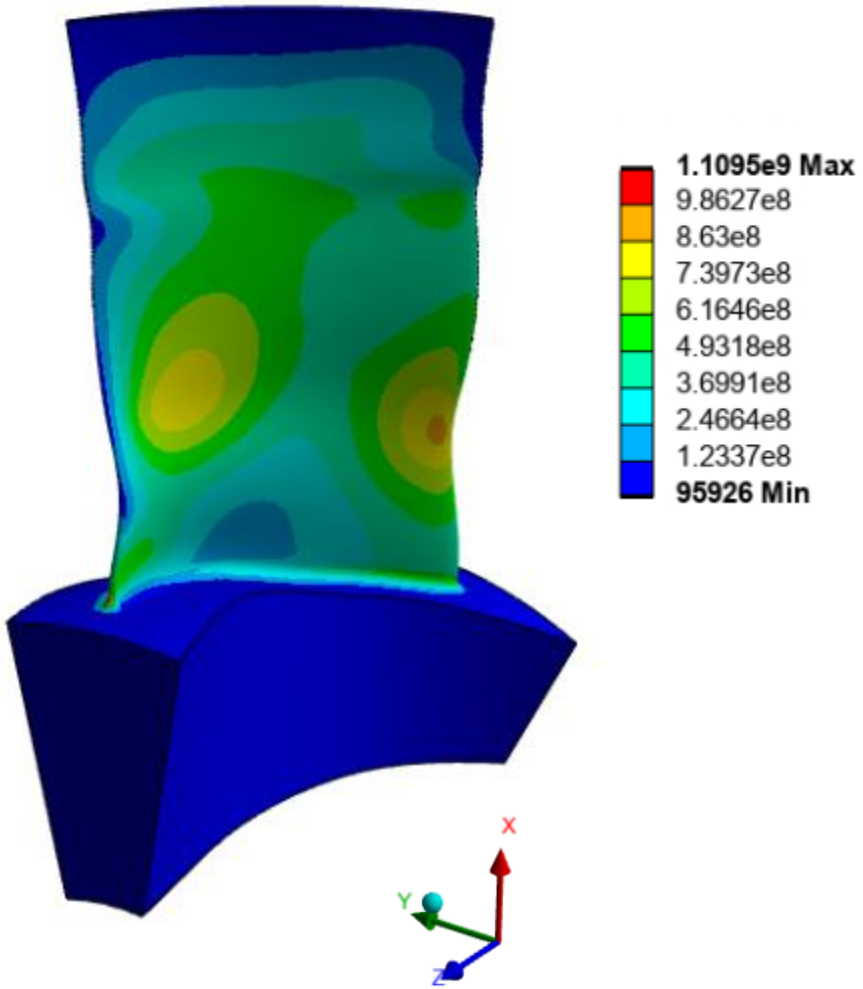

Trial Progression: 4 DVs
Margin of Safety
Trial Number
0
100
200
300
400
500
0
1
2
3
4


Trial Progression: 6 DVs
Margin of Safety
Trial Number
0
100
200
300
400
500
−1
0
1
2
3
4
5


Trial Progression: 8 DVs
Margin of Safety
Trial Number
0
100
200
300
400
500
−1
0
1
2
3
4
5

Baseline vs Optimized: Von-mises Stress Contours

Baseline Design



Optimized Design (6 DVs)



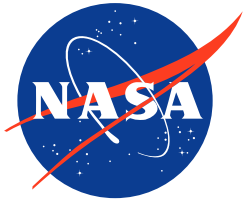
Applications and Future Work

- Bayesian optimization is currently being utilized by our group at NASA in the development of an open rotor design.
  - This approach is considering both the structural margin of safety and the aerodynamic performance in the objective function.
  - Current team members also include John Gillespie and Matthew Ha. We are guided by Mark Turner.

- An approach considering dynamic design spaces is also being researched.
  - This approach would allow the design space to dynamically shift as the optimization process progresses.
  - This approach is effective in cases where the optimal design is outside of the initial design space bounds.

- Alternative surrogate models are also an interest of future research.
  - Gaussian process surrogate models become much more expensive to generate as the amount of data increases.
  - After a certain threshold of data size, Bayesian Neural Networks may become a better choice for surrogate model.

Questions?



Glenn Research Center