Glenn Research Center

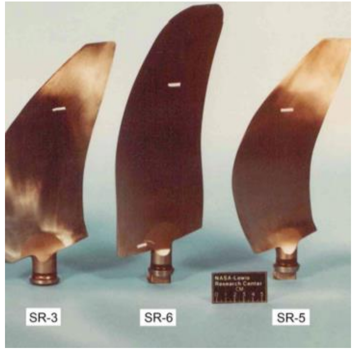
Structural Optimization of Turbomachinery Rotors: A Study of Machine Learning Surrogate Models

Kristopher Pierson[1], John Gillespie[1], Matthew Ha[2] and Joshua Stuckner[1]

July 23rd, 2025

1: NASA Glenn Research Center

2: HX5, LLC, NASA Glenn Research Center



This work was supported by:

NASA AATT Project

AATT: Advanced Air Transport Technology

Background and Description of Study

- Previous researchers, including this group, have utilized machine learning surrogate models for design optimization. Some examples are included in the text.
- In this work, two different types of machine learning models are evaluated for their ability to represent design spaces on a range of training samples and design variables.
  - Multi-layer perceptron (MLP), a type of neural networks
    - Three configurations are studied: 1, 3 or 5 hidden layers.
  - Random Forest Regressor
    - 100 decision trees for each model.
    - An informal investigation of increasing the number of decision trees was performed, but no benefit was observed.
- A python package called scikit-learn is used as the basis for these two models.
- Three similar design spaces are studied, each with a different number of design variables (DVs): 4 DVs, 6 DVs and 8 DVs.
- Each design space is sampled using Latin hypercube sampling (LHS) of various sizes.
- Finally, the best performing surrogate models of each type are used for design optimization.

Sampling and Training Methodology

- LHS training set sizes of 500, 1000, 2000 and 5000 are used to train the models.
  - A corresponding LHS validation set is generated for each training set that is 20% the size of its associated training set.
- Blade geometry is generated using T-Blade3, 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. This is never used in any portion of model training. It is used exclusively to evaluate the models.
- 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.

Step 1: Choose a Design Space

↓

Step 2: Sample the Space with LHS

↓

Step 3: Evaluate Each Point in the Sample with FEA

↓

Step 4: Train a Machine Learning Model with the Data

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

The Design Spaces

- Three design spaces are studied. The design variables in each design space are lean and sweep control points.
  - The number of spanwise control points differs between the design spaces.
- The 4 DV space contains control points at 50% and 100% span locations.
- The 6 DV space: 25%, 50 and 100% span locations
- The 8 DV space: 25%, 50%, 75% and 100%.
- 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

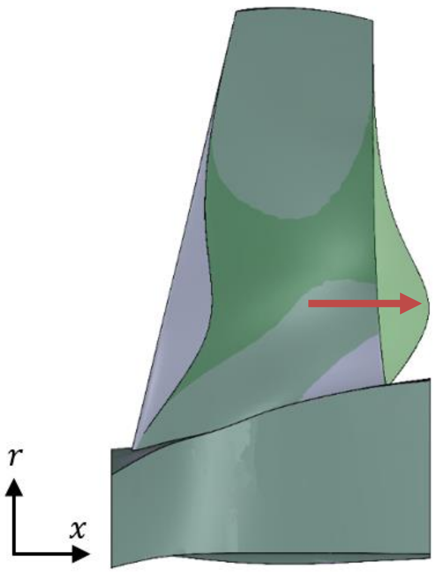
Variable	Range
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

Design Space 2

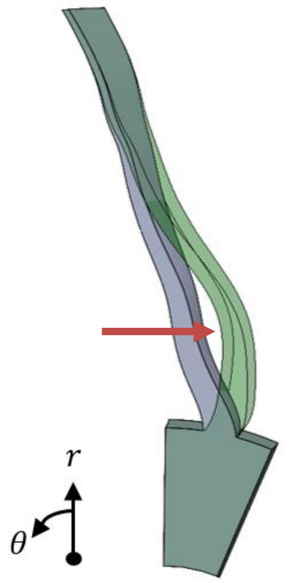
Variable	Range
$\Delta m'_{25}$	(-0.17,0.03)
$\Delta m'_{50}$	(-0.17,0.03)
$\Delta m'_{100}$	(-0.17,0.03)
$\Delta\theta_{25}$	(-0.17,0.03)
$\Delta\theta_{50}$	(-0.17,0.03)
$\Delta\theta_{100}$	(-0.17,0.03)

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



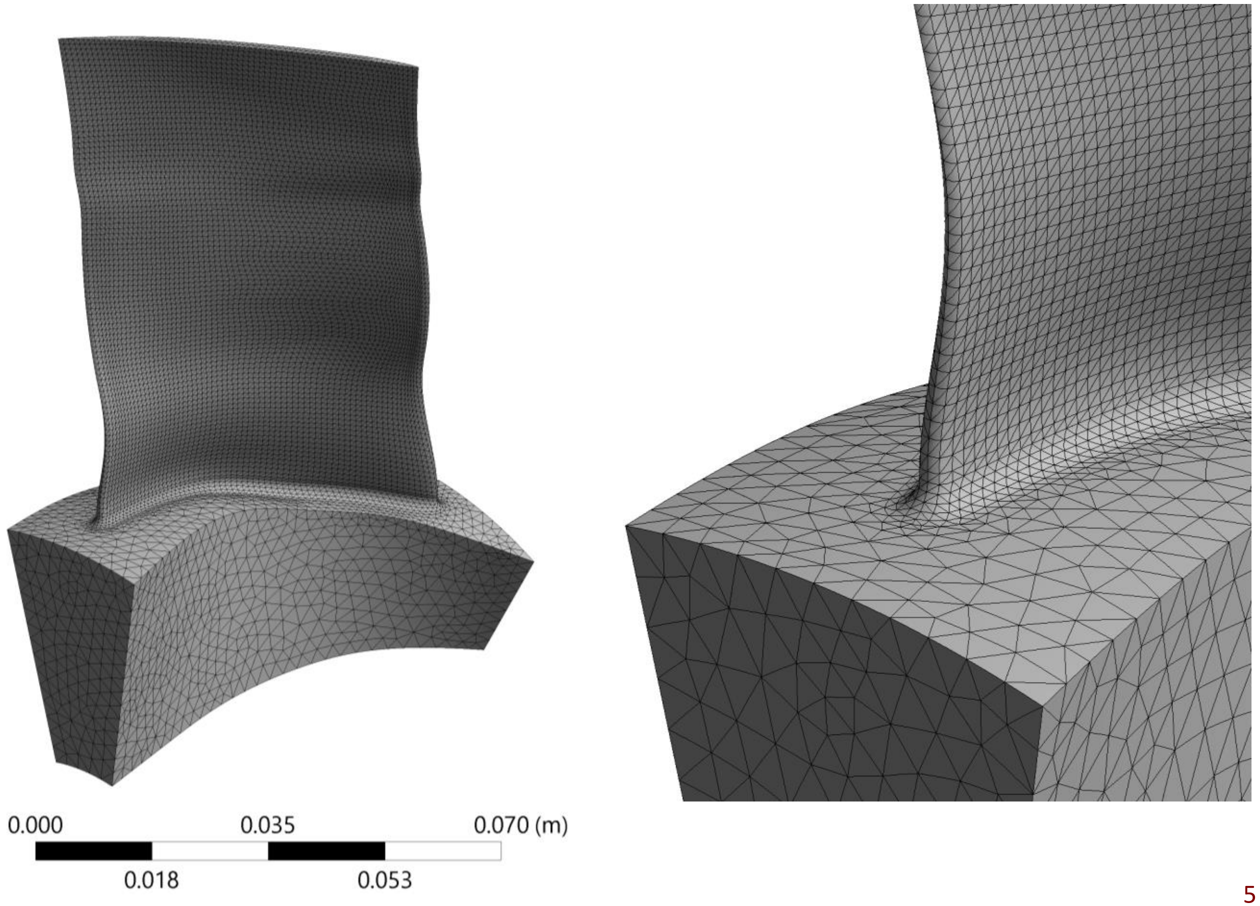
Lean ($\Delta\theta_{25}$)

Baseline Model

- The baseline model is the tail cone thruster 7-28-8.
  - Designed by UC students
  - Openly available on T-Blade3 github page, which is referenced in the paper.
- All lean and sweep control points are set as 0 for the baseline design.
- Titanium Ti-6Al-4V 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.

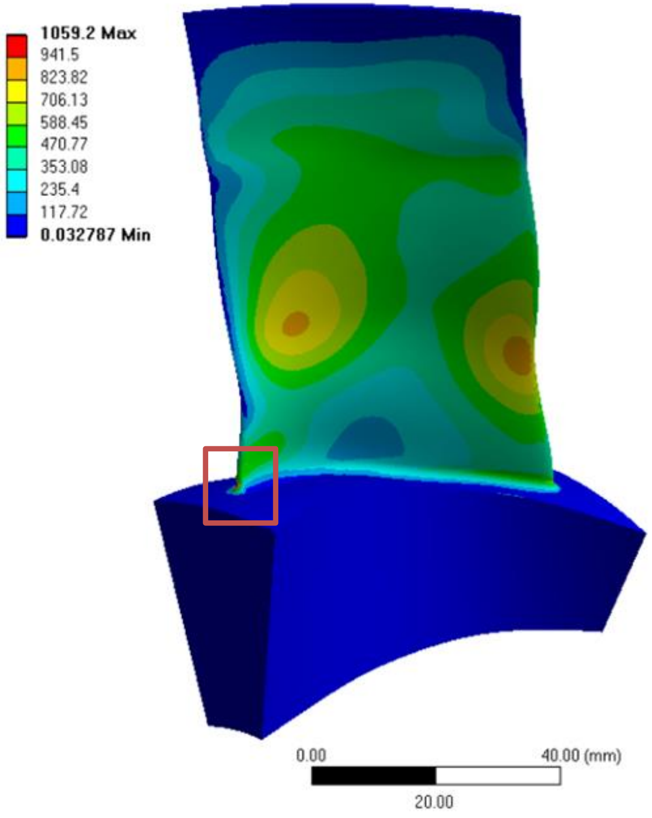
Material Properties of Ti-6Al-4V (Grade 5)

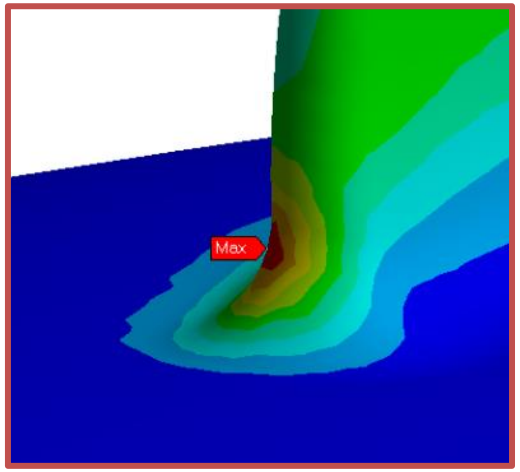
Density [kg/m^3]	4430
Young's Modulus [Pa]	1.14E+11
Poisson's Ratio	0.33
Tensile Yield Strength [Pa]	1.10E+09



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.

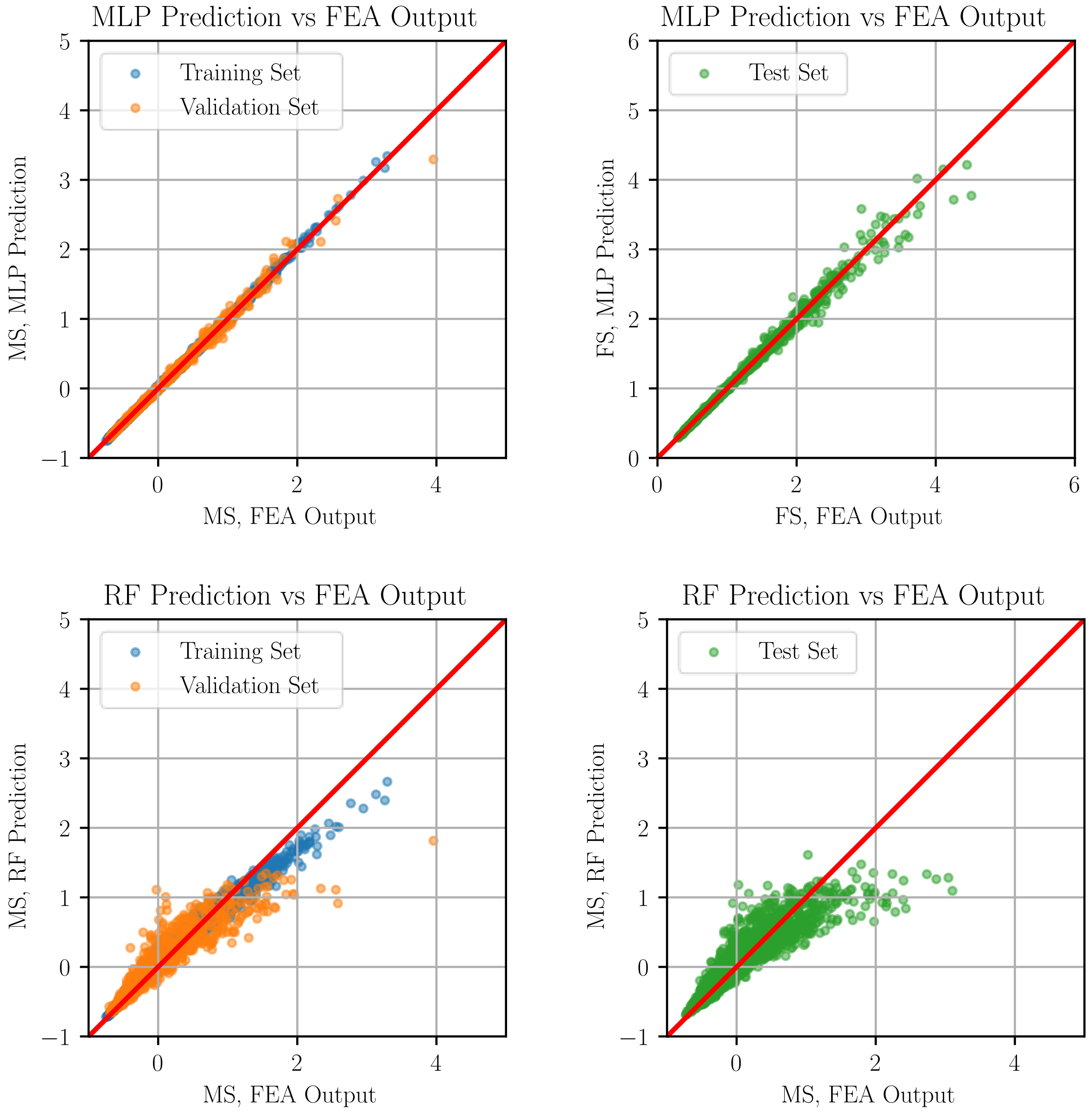




Accuracy Evaluation Metric

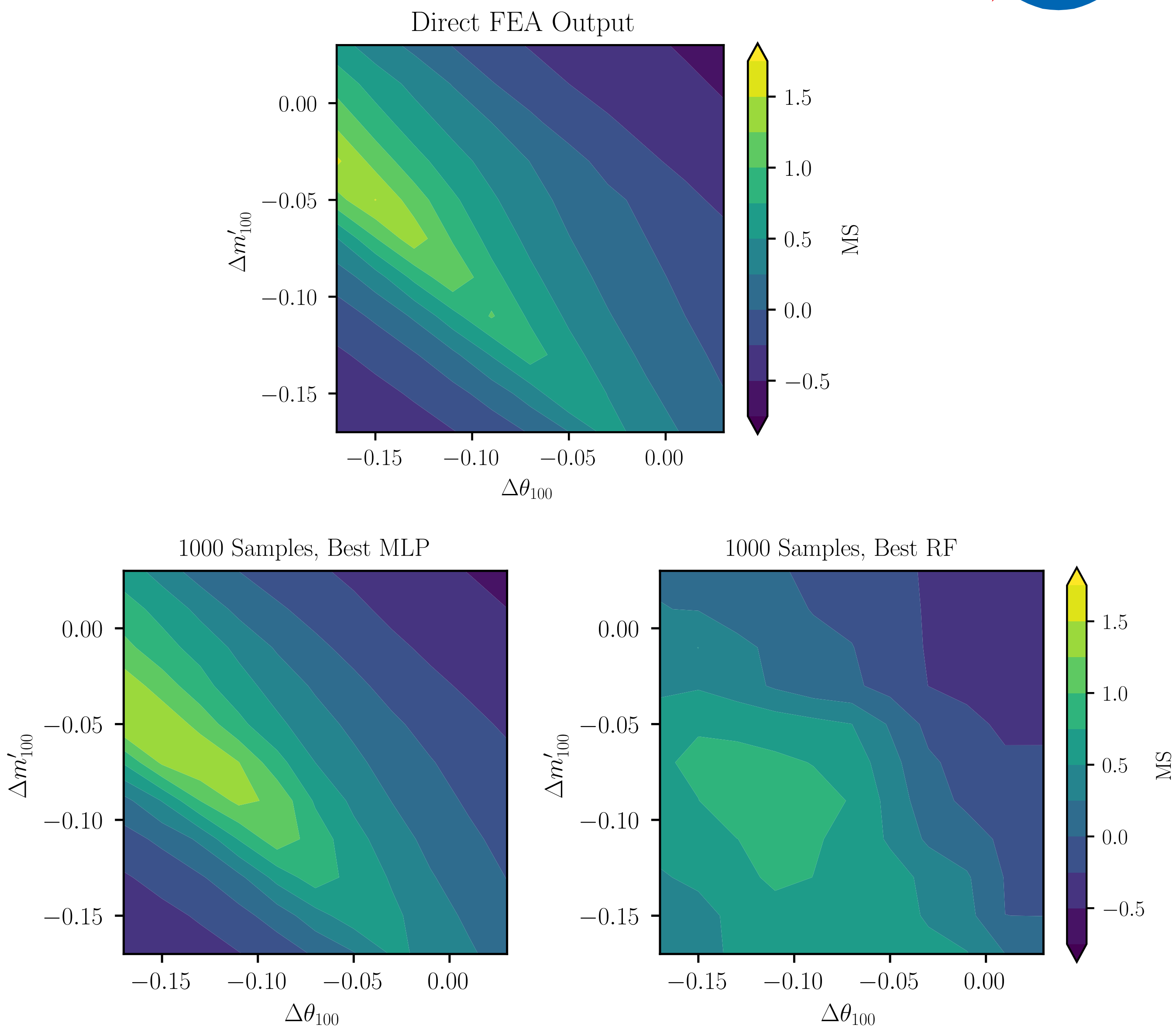
- Models are initially evaluated using the coefficient of determination (R^2) metric that is often used to assess the accuracy of trendlines.
- The plots to the right show MS values predicted by the surrogate model compared to MS values from the FEA.
  - The x-axis on the plots to the right represents the FEA result
  - The y-axis is the 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.

	Best MLP	Best 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



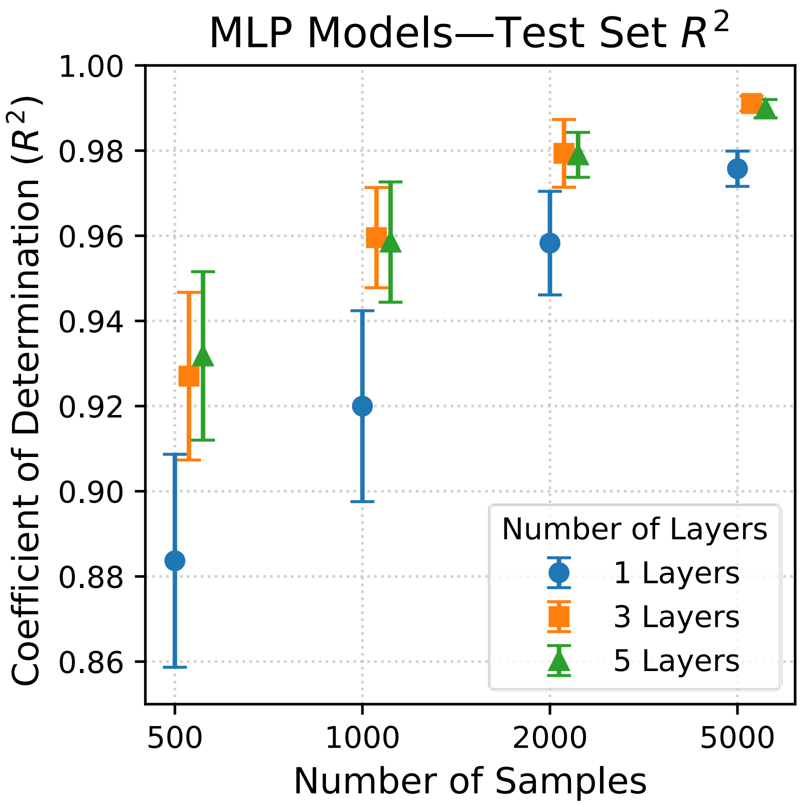
6 DV Surrogate Model Visualization

- The visualization to the right reinforces the higher accuracy of the MLP models compared to the RF models.
- The plots are a slice of the six dimensional design space.
  - To generate the slice, $\Delta m'_{100}$ and $\Delta\theta_{100}$ are varied across the whole range of the design space while other variables are fixed at the center of the design space.
- The plots shown on the right are for the best performing MLP model and RF model trained with 1000 samples.
- Visual inspections shows that the MLP model represents the real response much more accurately compared to the RF model.
  - Given the result of the R^2 evaluation, this is consistent with the expected behavior.
- In the interest of time, RF model evaluation will not be shown for the rest of this presentation, but it can be found in the paper.



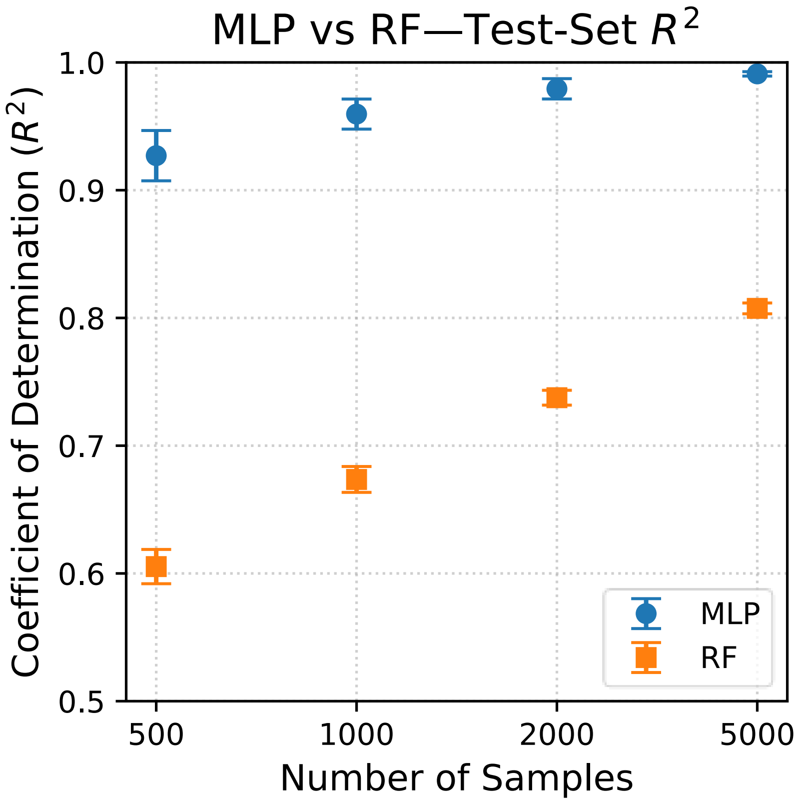
6 DV Design Space - MLP Evaluation

- 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 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 on the basis of the validation fraction.
- The performance of the three layer MLP was significantly better than the one layer model.
  - However, no significant performance improvement consistently occurred when moving from three to five layers.



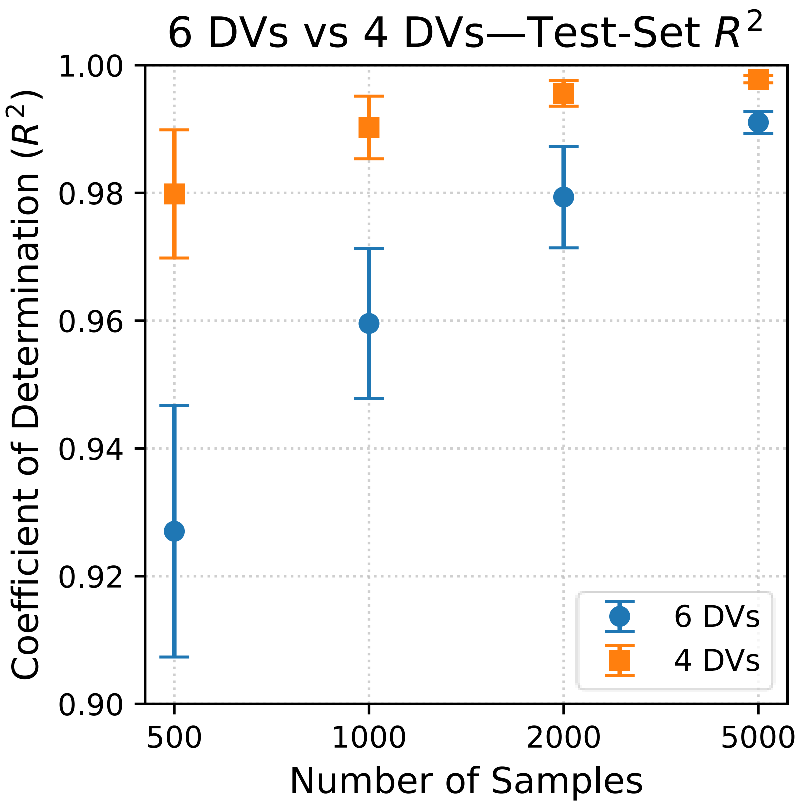
6 DV Design Space - Random Forest

- In terms of R^2 value, the RF models performed much more poorly compared to the neural network models.
- The best performing RF model trained with 5000 samples had poorer accuracy compared to the poorest performing MLP model trained with only 500 sample.
  - This is consistent across all three design spaces studied n this work.



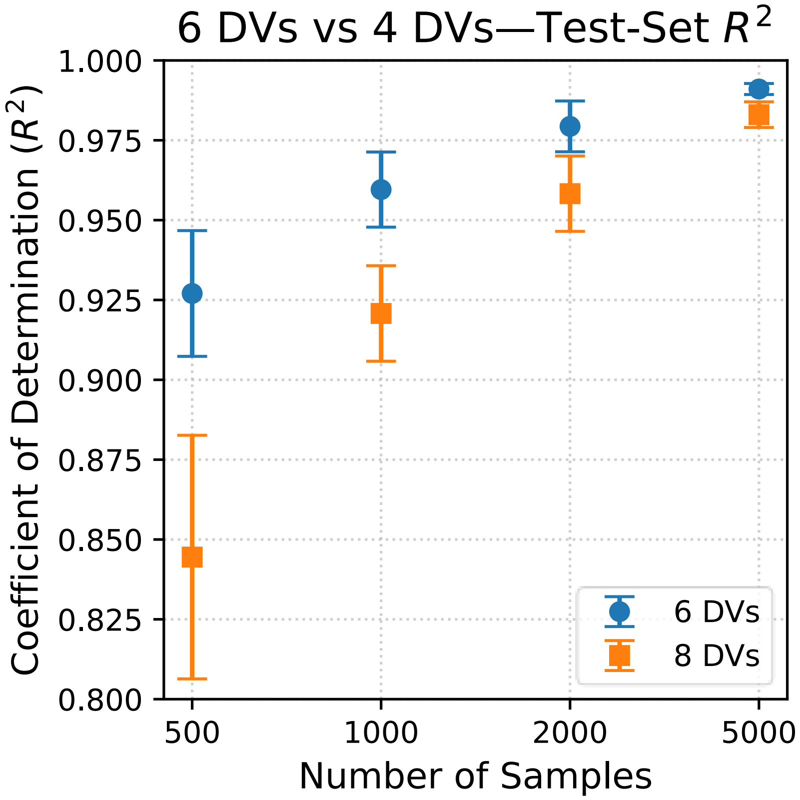
4 DV MLP Evaluation

- The figures shown to the right represent the three layer MLP models.
- As expected, the surrogate models for the 4 DV space attain higher accuracy when trained with an equal number of samples as the 6 DV case.
- This was more significant for the models trained with 500 samples.
- The effect was less pronounced for the models trained with 5000 samples.
  - For the large number of training samples, the 6 DV models still generally achieve an R^2 greater than 0.99 indicated a high accuracy in the representation of the design space.
- 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, the only difference is the random seed.
- Not shown in the figures to the right: the accuracy is again seen to improve when moving from one to three layers, but no significant improvement is seen when moving from three to five layers.



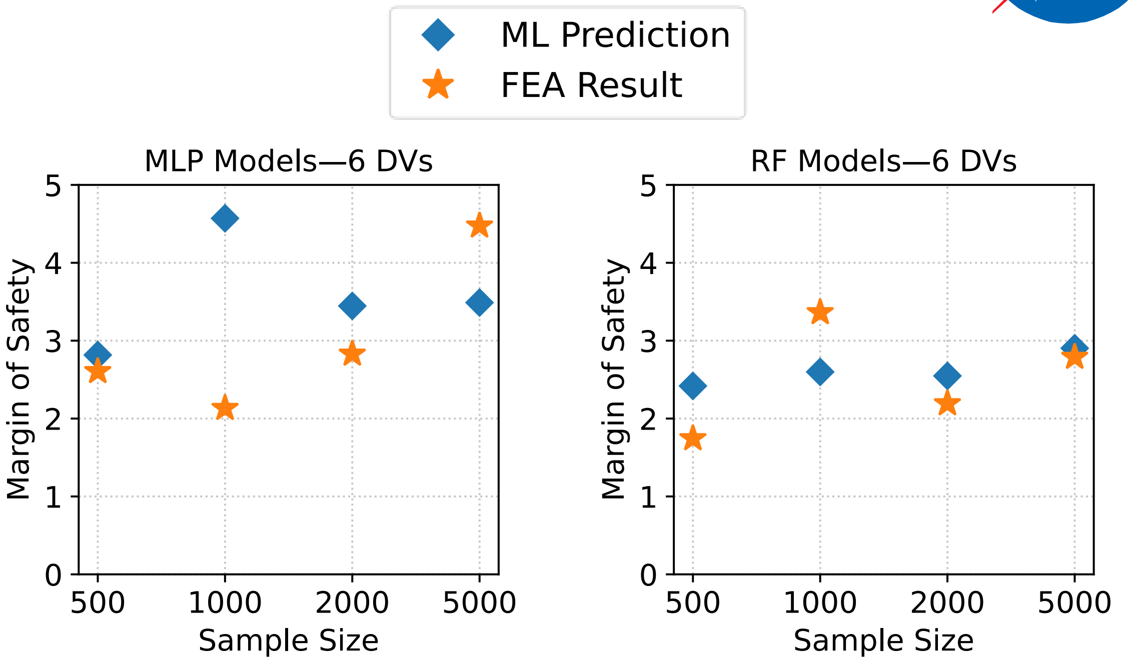
8 DV MLP Evaluation

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



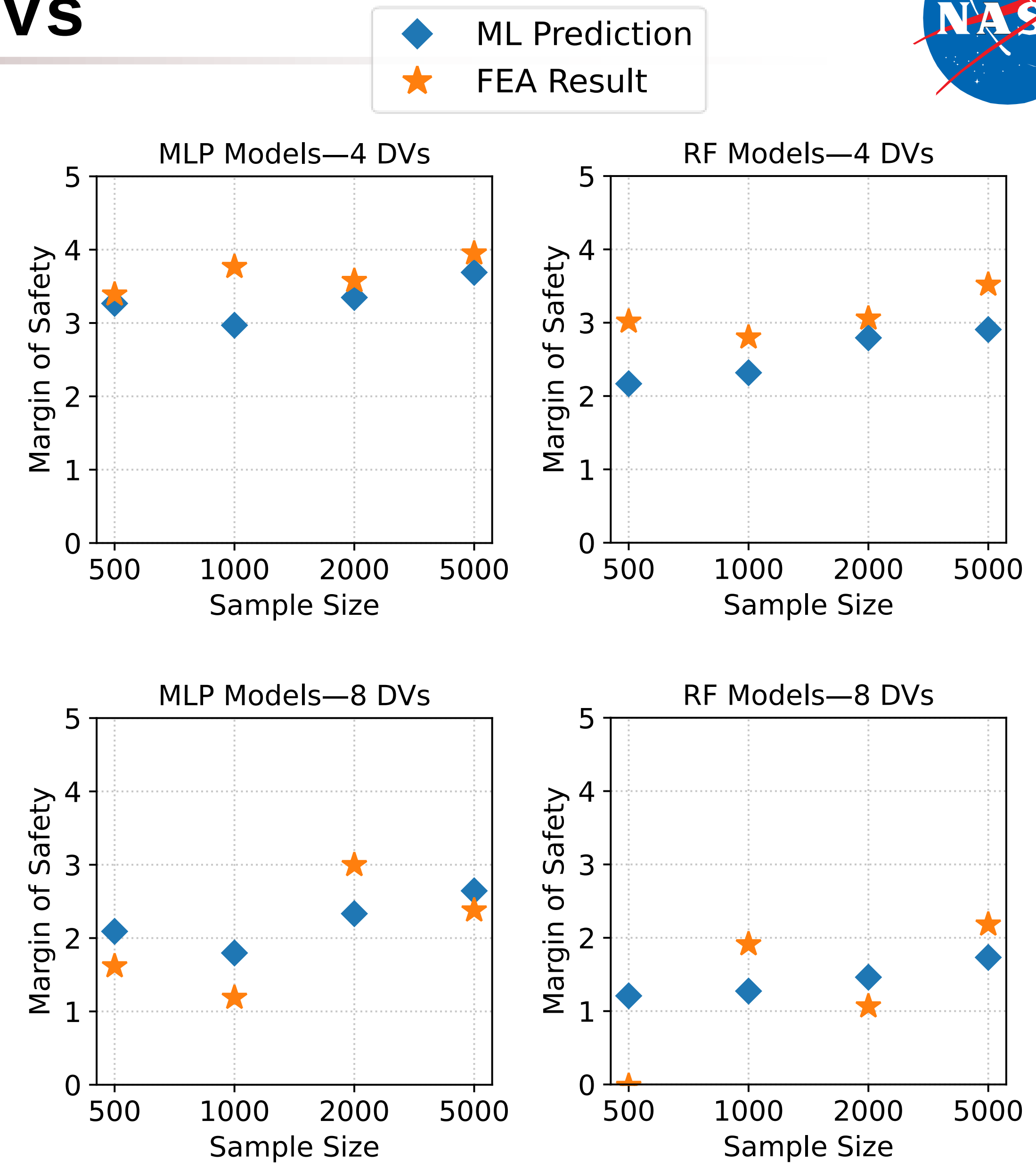
Design Optimization, 6 DVs

- The performance of the surrogate models in design optimization provides a different look at the performance of the surrogate models.
  - 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.
  - This is available in the Python package SciPy.
  - Default parameters are used
- The optimization is run 100 times
  - Differential Evolution has some inherently random aspects, so each result could be different.
  - Most of the optimization runs resulted in the same design while some were stuck in a local minimum.
- 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
- However, they were unable to achieve the performance of the best result from 6 DVs.
  - This may be a limitation of having too few design variables.
- 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, which is not surprising given the results of the accuracy evaluation using the R^2 metric.



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

- 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 was particularly apparent in the design optimization portion of the work

- As a next step in this research, focus will be shifted to Bayesian Optimization methods, which will avoid spending excessive resources in regions of poor performance.

Questions?

Glenn Research Center

Extra Slides

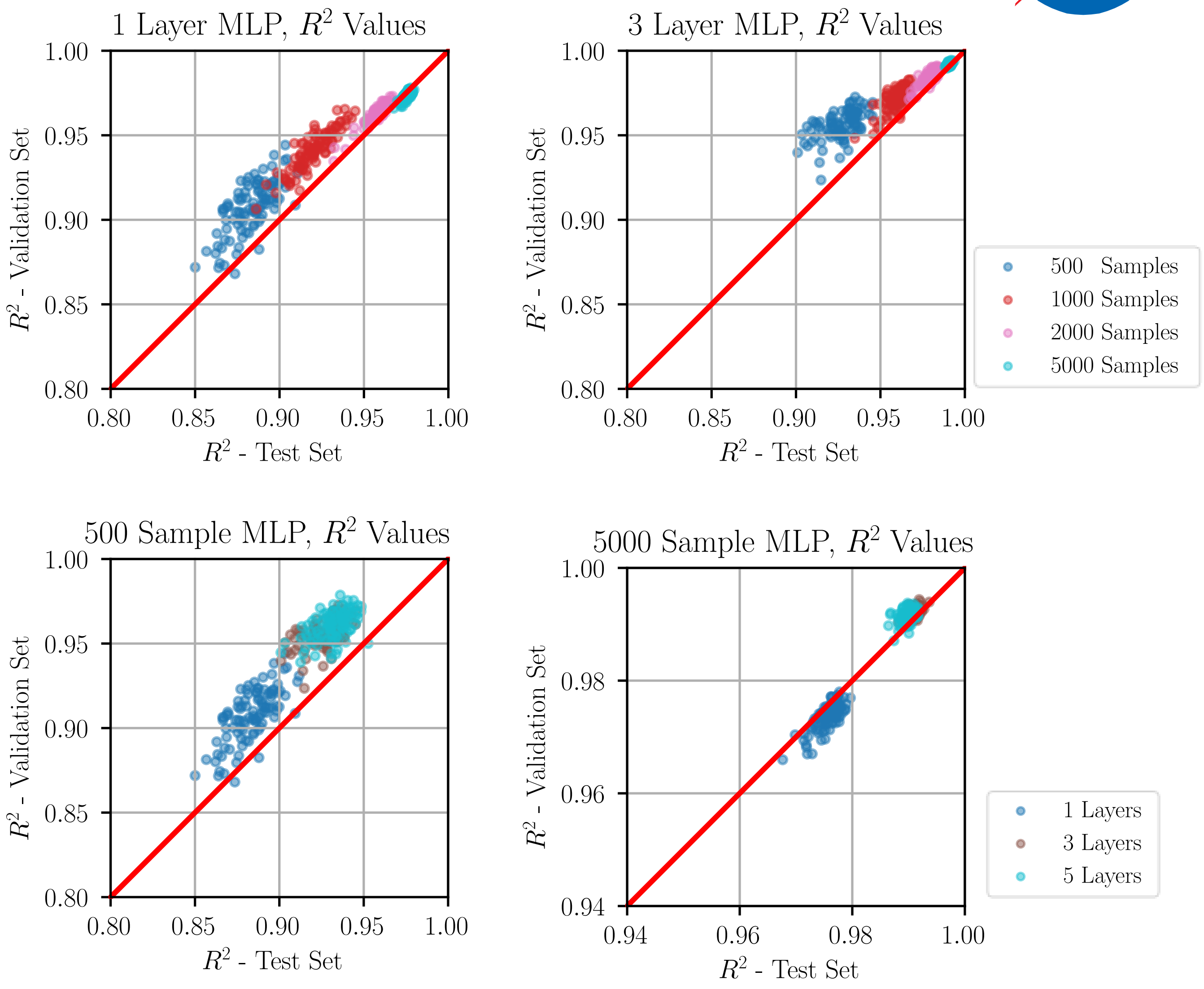
Glenn Research Center

Coefficient of Determination

- $R^2 = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - \bar{y})^2}$

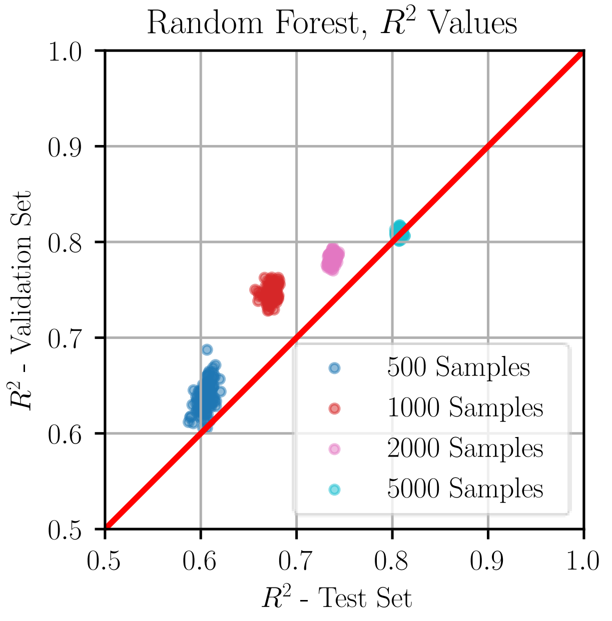
6 DV Design Space - MLP Evaluation

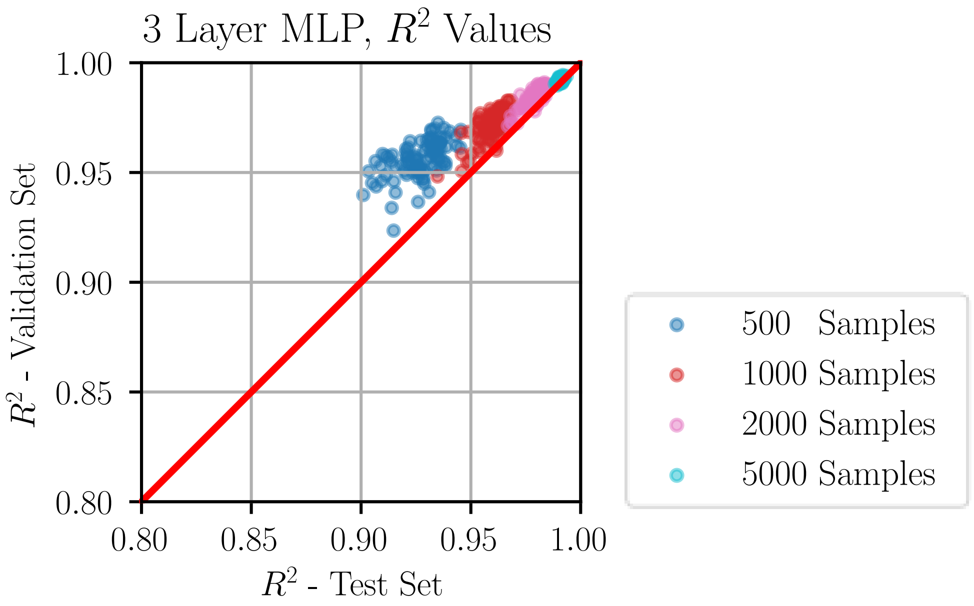
- The red line on these plots does not have the same meaning as on the plots from the previous page. It is simply used as a visual aid to determine if performance is better on the test or validation set.
- Each type of machine learning model is trained 100 times using 100 different random seeds.
  - Neural network model training has inherent randomness built into it through weight initialization and data shuffling.
- The 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 on the basis of the validation fraction.
- The performance of the three layer MLP was significantly better than the one layer model.
  - However, no significant performance improvement consistently occurred when moving from three to five layers.



6 DV Design Space - Random Forest

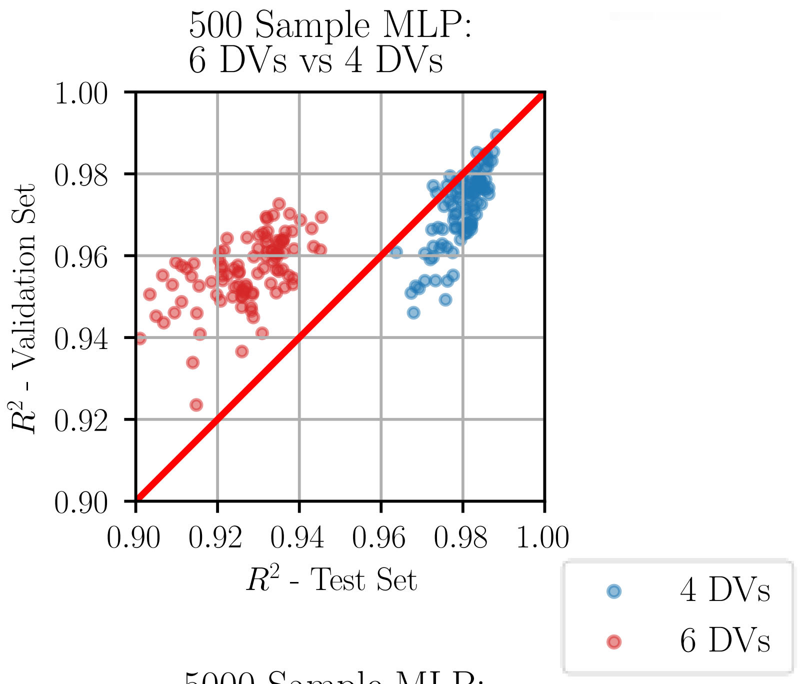
- In terms of R^2 value, the RF models performed much more poorly compared to the neural network models.
  - Please note that the x and y axis scales are different in the two plots to the right.
- The best performing RF model trained with 5000 samples had poorer accuracy compared to the poorest performing MLP model trained with only 500 sample.
  - This is consistent across all three design spaces studied n this work.

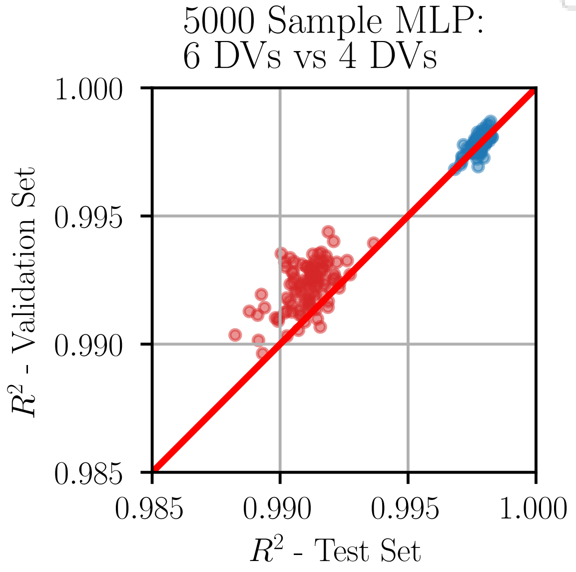




4 DV Neural Network Evaluation

- The figures shown to the right represent the three layer MLP models.
- As expected, the surrogate models for the 4 DV space attain higher accuracy when trained with an equal number of samples as the 6 DV case.
- This was more significant for the models trained with 500 samples.
- The effect was less pronounced for the models trained with 5000 samples.
  - For the large number of training samples, the 6 DV models still generally achieve an R^2 greater than 0.99 indicated a high accuracy in the representation of the design space.
- 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, the only difference is the random seed.
- Not shown in the figures to the right: the accuracy is again seen to improve when moving from one to three layers, but no significant improvement is seen when moving from three to five layers.





8 DV Neural Network Evaluation

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