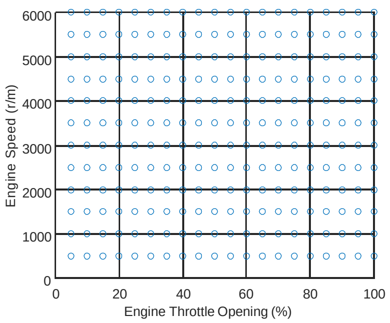
MathWorks
AUTOMOTIVE
CONFERENCE 2023
North America

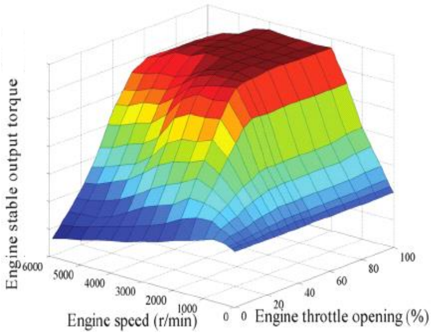
Adaptive Design of Experiment for Simultaneous Modeling and Optimization with Artificial Intelligence

Yan Wang, Ford Motor Company

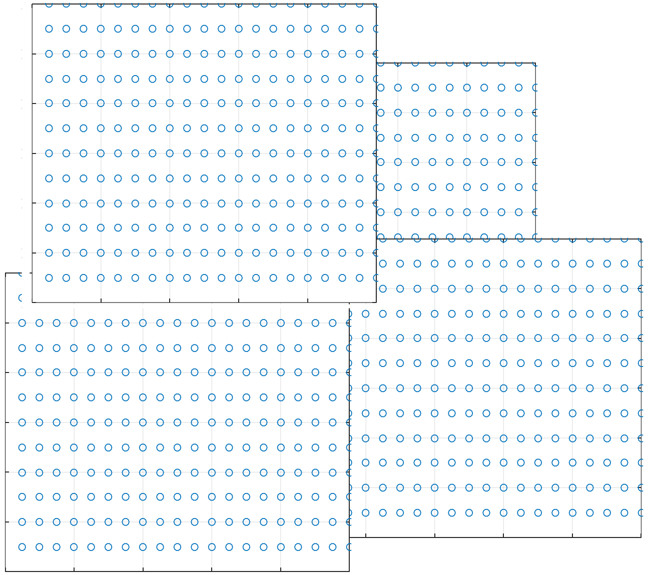
MathWorks

Why are we looking at this

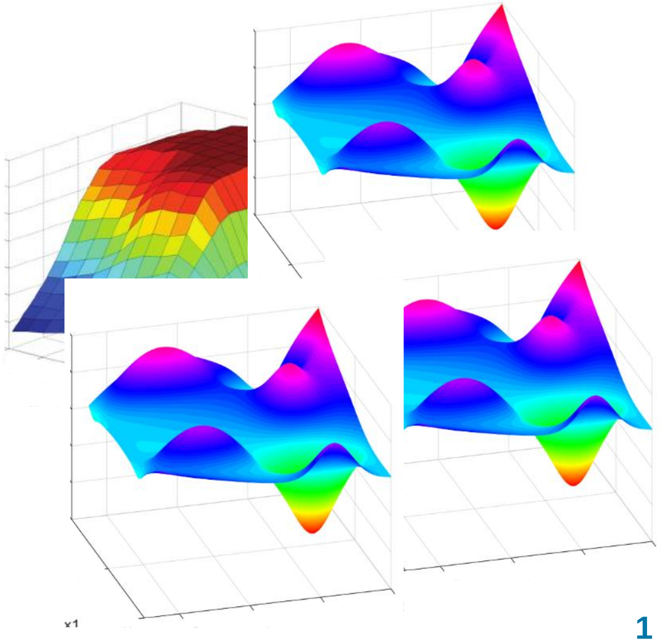




Characterization & Calibration

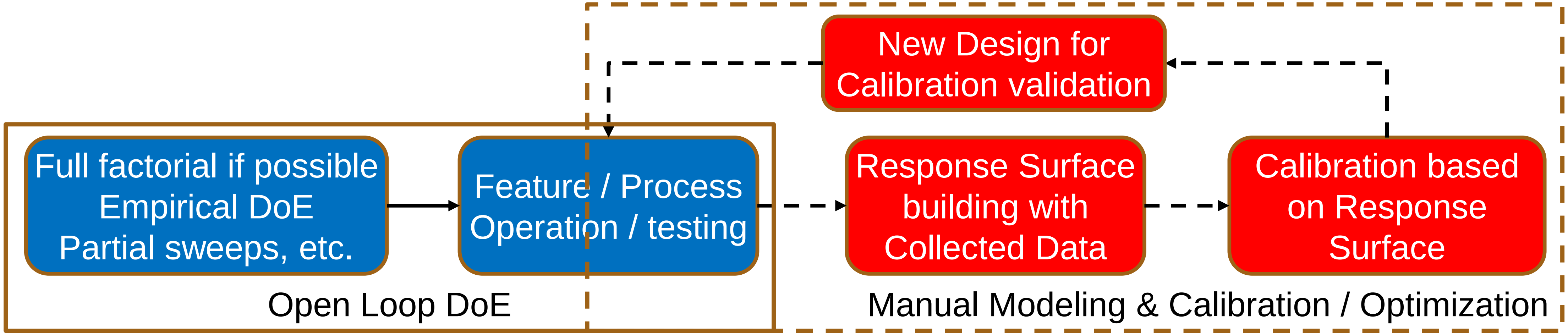


- Sweep based DoE fits well to generate data for characterization and calibration
- However, tests needed to characterize and calibrate the system increase exponentially with the number of design variables
- In addition, extra degree of freedom may introduce more nonlinear responses



It's not sustainable to run sweep based DoE for high DoF characterization

Current Design of Experiment Practice: Open loop + Manual



Full factorial if possible
Empirical DoE
Partial sweeps, etc.

Feature / Process
Operation / testing

Open Loop DoE

New Design for
Calibration validation

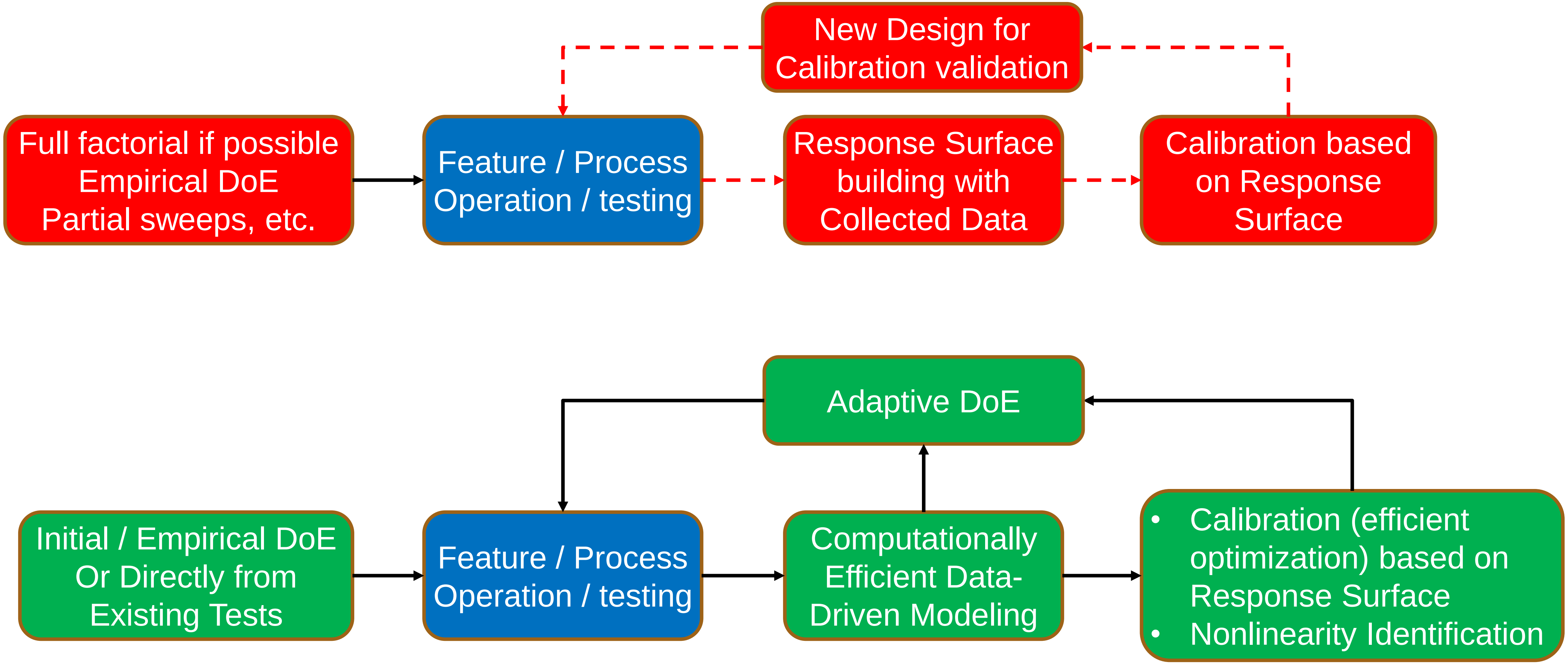
Response Surface
building with
Collected Data

Calibration based
on Response
Surface

Manual Modeling & Calibration / Optimization

- Open loop DoE focuses on design space, and doesn’t take into account the response surface
  - May miss nonlinearity or have too many tests in linear regions
  - Doesn’t necessarily test / validate the optimal region
- Manual modeling and calibration need human intervention
  - Build response surface / model with data collected from open loop DoE
  - Optimal calibration based on the response surface may be in the regions that don’t have sufficient data / resolution, or may even be found with extrapolation, and therefore, may not be duplicated on hardware
  - Likely need multiple human-in-the-loop iterations

Adaptive Design of Experiment Practice: Closed loop + Automatic



New Design for Calibration validation

Full factorial if possible
Empirical DoE
Partial sweeps, etc.

Feature / Process Operation / testing

Response Surface building with Collected Data

Calibration based on Response Surface

Adaptive DoE

Initial / Empirical DoE
Or Directly from Existing Tests

Feature / Process Operation / testing

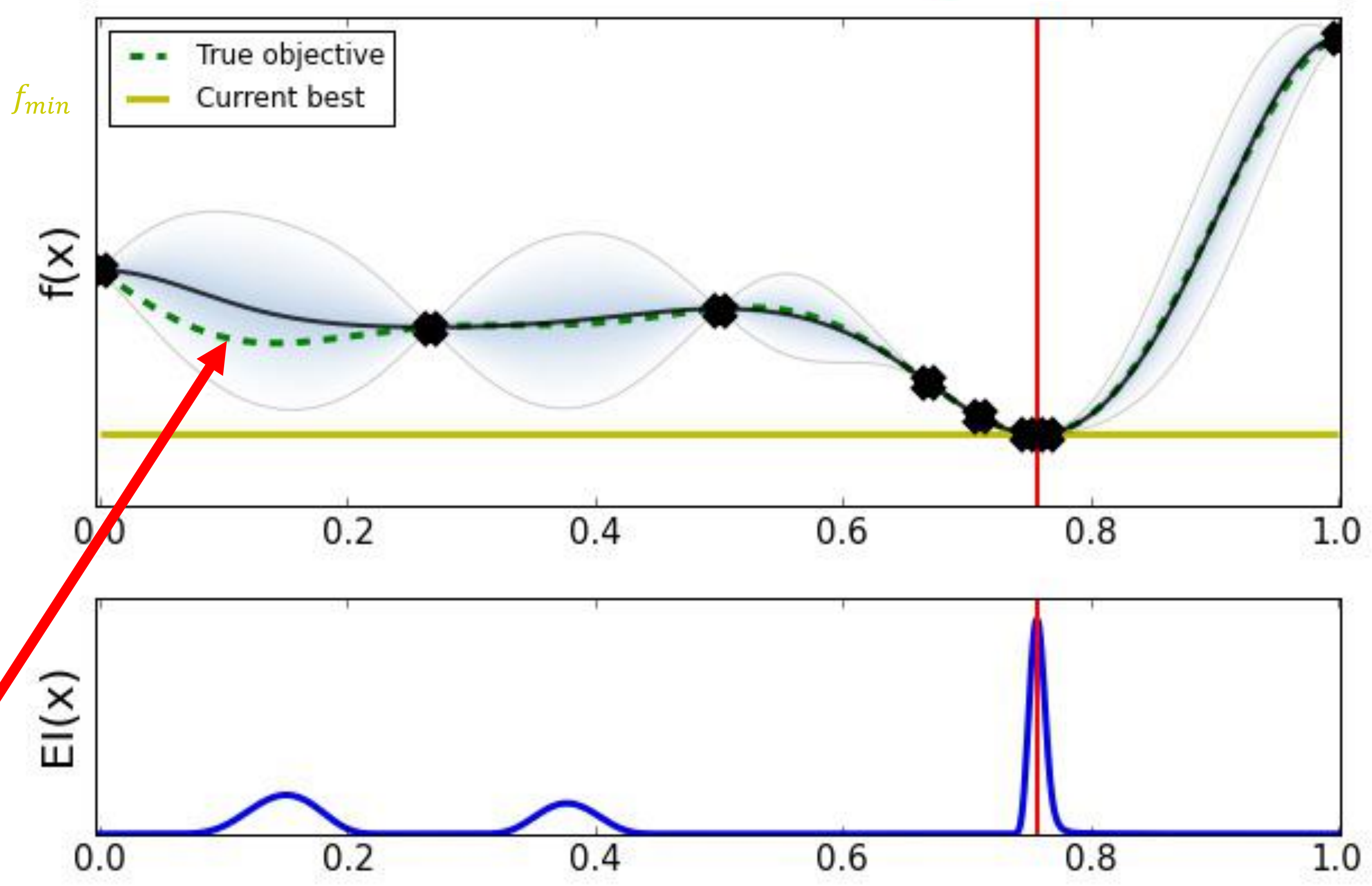
Computationally Efficient Data-Driven Modeling

- Calibration (efficient optimization) based on Response Surface
- Nonlinearity Identification

Key enabler is efficient algorithms (modeling and optimization) to realize **online** implementation

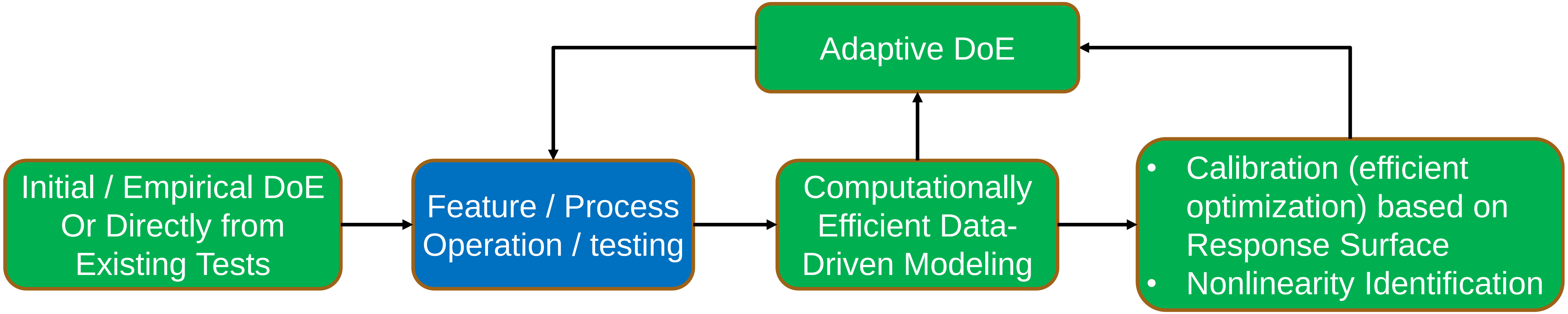
Adaptive DoE based on Bayesian Optimization Concept

- Bayesian optimization is a sequential design strategy for **global optimization** of black-box functions and is usually employed to optimize **expensive-to-evaluate** functions.
- Calculate acquisition function to determine where to evaluate the function next to **achieve optimality**, considering both mean and variance
- Extending the concept to cover the nonlinearity serves the adaptive DoE purposes



©Javier Gonzalez 2017

Simultaneous Modeling and Calibration



Adaptive DoE

Initial / Empirical DoE Or Directly from Existing Tests

Feature / Process Operation / testing

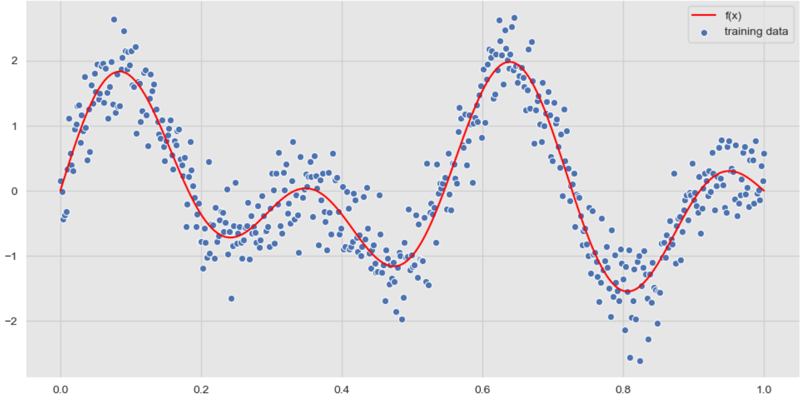
Computationally Efficient Data-Driven Modeling

- Calibration (efficient optimization) based on Response Surface
- Nonlinearity Identification

- Adaptive /online DOE builds response surfaces (modeling) while running optimization (calibration)
  - Add inputs **and outputs** data-driven algorithm to explore design space
    - Build robust surrogate model online, with **small but sufficient amount of data**
    - Identify nonlinear regions
    - Online optimization
  - New designs based on both optimal and nonlinear regions identification from surrogate model
- Modeling achieved with converged surrogate models
- Calibration achieved with optimization and subsequent validation (part of DoE)

Data-Driven Model

GPR



$$y \sim GP(m(x), k(x,x') + \delta_{ij}\sigma_n^2)$$

$$k(x,x') = \sigma_f^2\ exp(-\frac{1}{2\ell^2}||x-x'||^2)$$

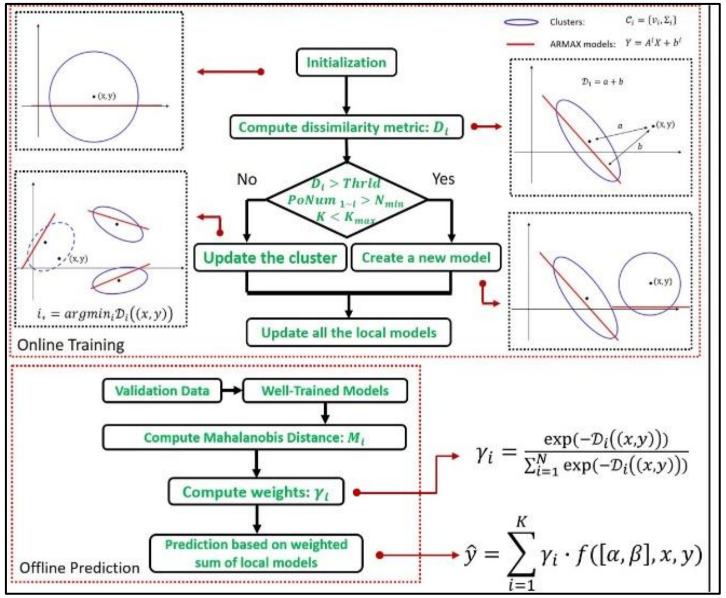
$$f^*|X,y,X^* \sim \mathcal{N}(\bar{f}^*, \Sigma^*)$$

$$\bar{f}^* = \mu^* + K(X^*,X)[K(X,X)+\sigma_n^2 I]^{-1}(y-\mu)$$

$$\Sigma^* = K(X^*,X^*) - K(X^*,X)[K(X,X)+\sigma_n^2 I]^{-1}K(X,X^*)$$

	Fitting Rate	**Time (sec)**	
	%	Training	Prediction
GPR	99.07	32.74	6.55
STF	94.87	0.73	1.9
STF with GA	98.39	10.32	1.8
Incremental STF	98.88	0.45	1.85

STF

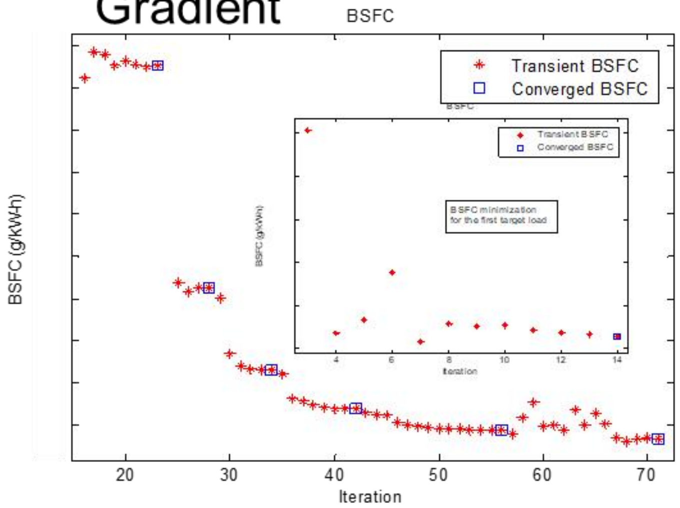


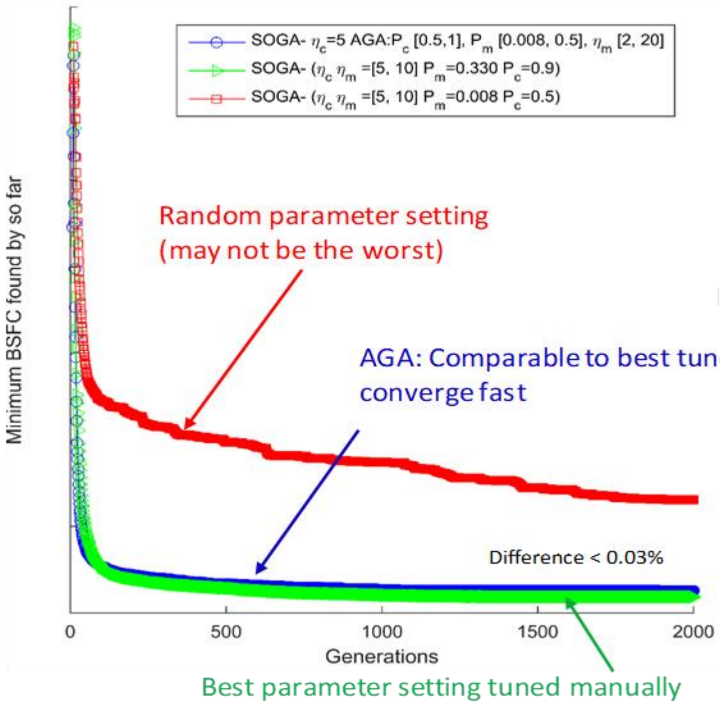
- Gaussian Process Regression (GPR) models are good candidate to deal with measurement variance from any physical system
- Spatial Temporal Filter (STF) model as an internal MATLAB based tool has demonstrated comparable performance but has the advantage of computation time (training & prediction)

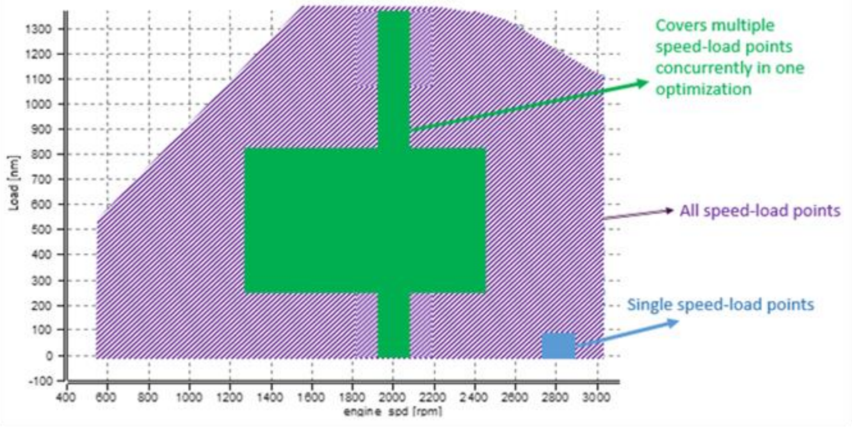
Key Challenge is to be efficient for online implementation

Online Optimization

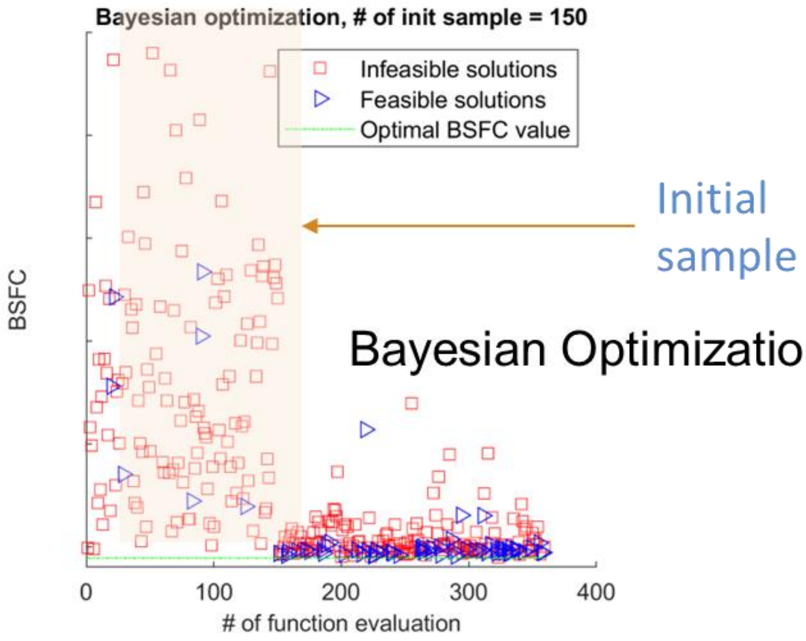
Gradient







Concurrent Optimization



Bayesian Optimization

- Different optimization schemes have been developed over the years
- Online optimization capability is the key!

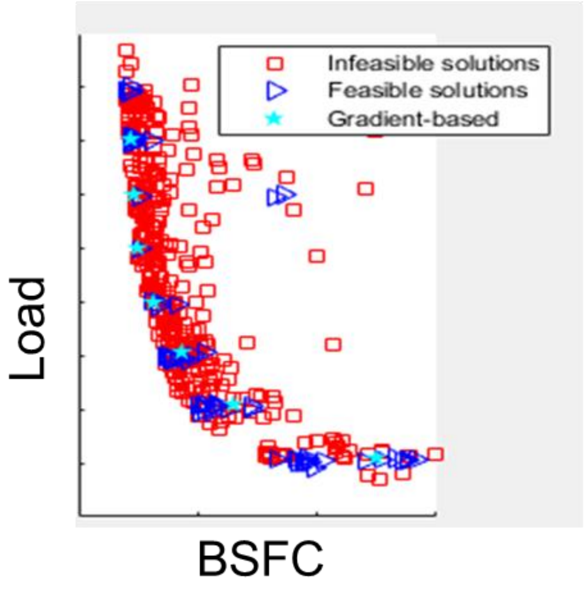
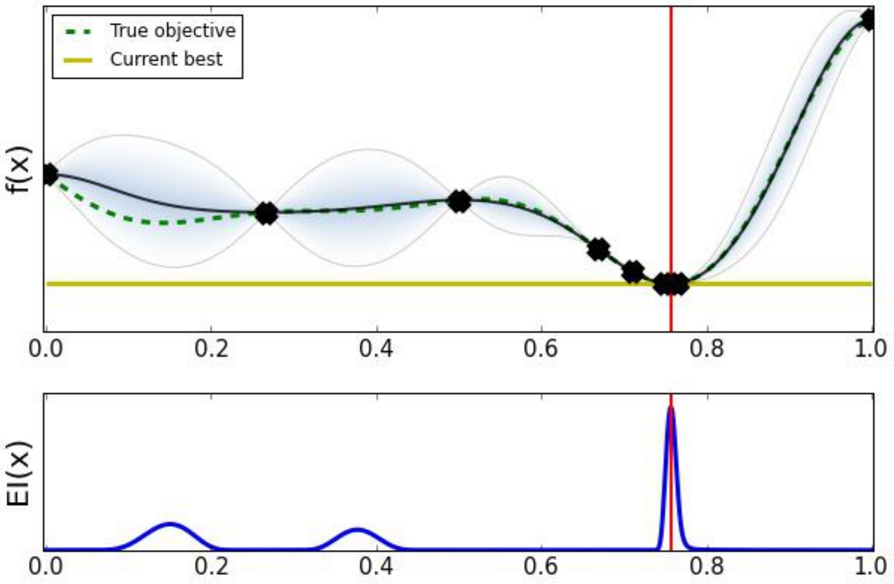
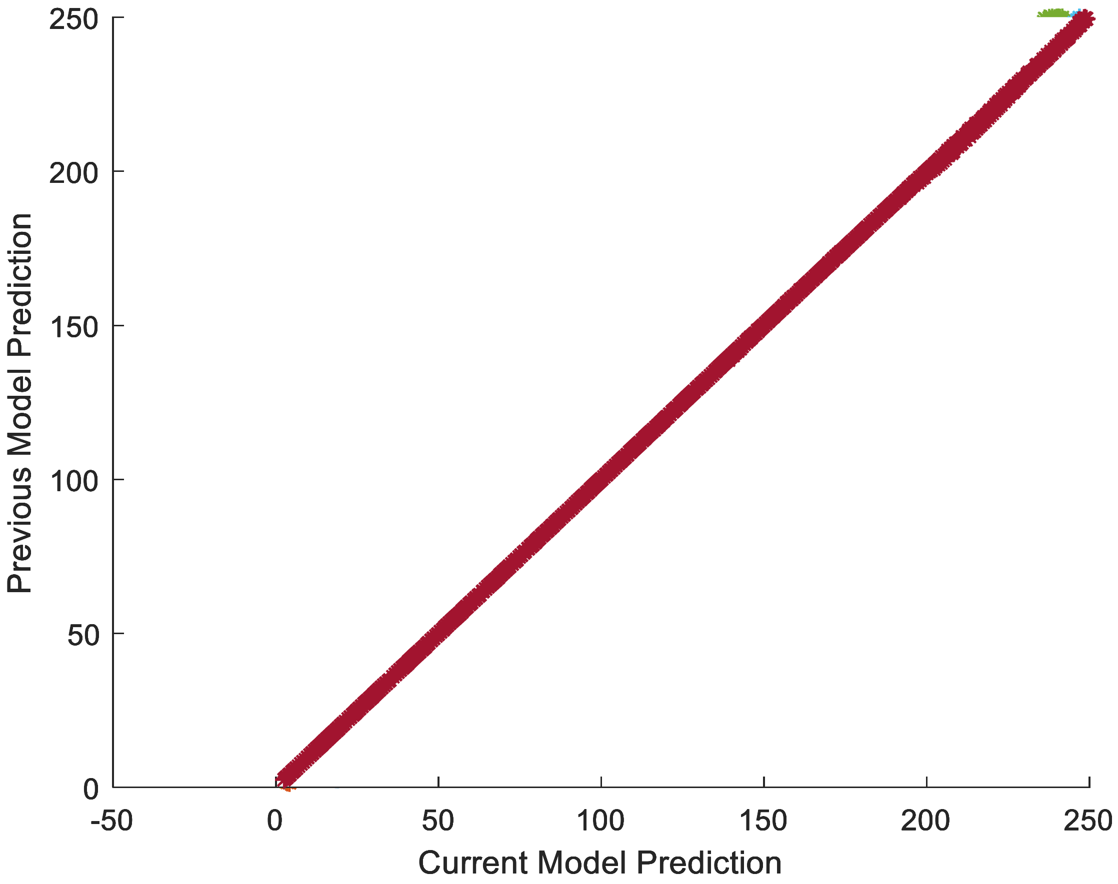


Illustration of Surrogate Models For Nonlinearity Identification

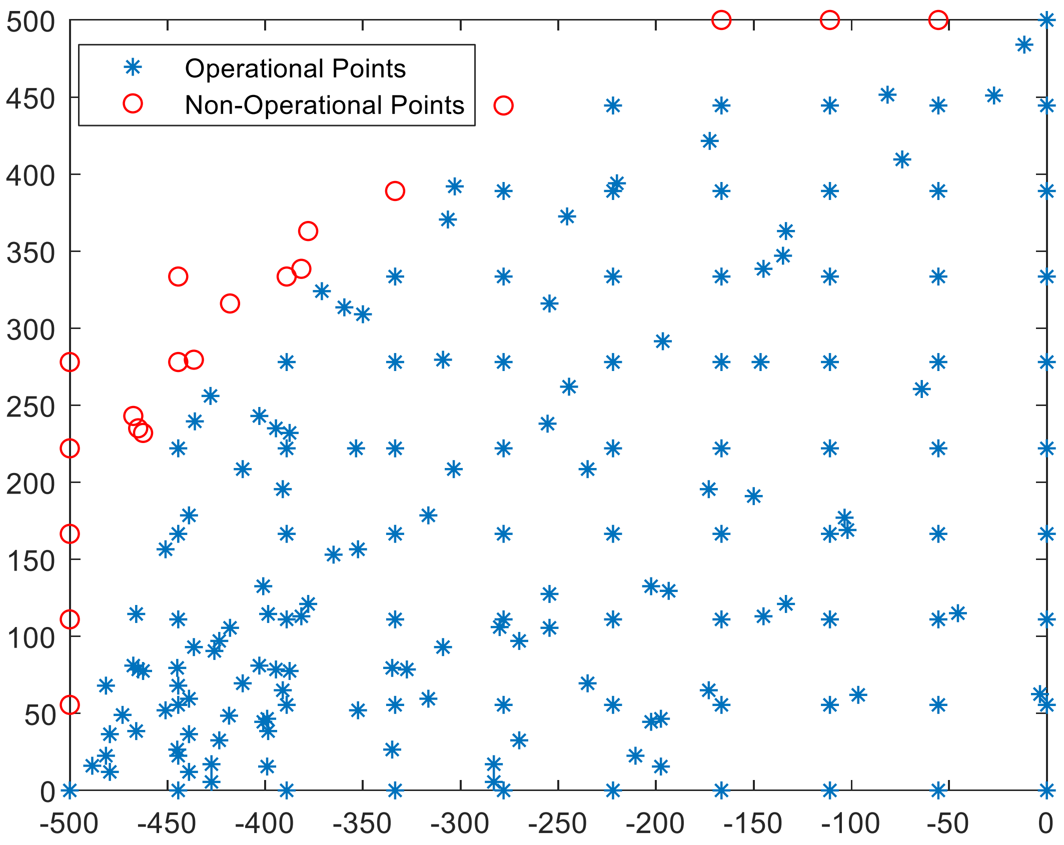


- Difference between consecutive surrogate model iterations (different data lengths) point to the nonlinearity (curvatures) that dictates design to capture
- Example of a motor calibration problem



Multiple numerical examples (for lower dimension problems) have confirmed the convergence to “ground truth” when surrogate models converge

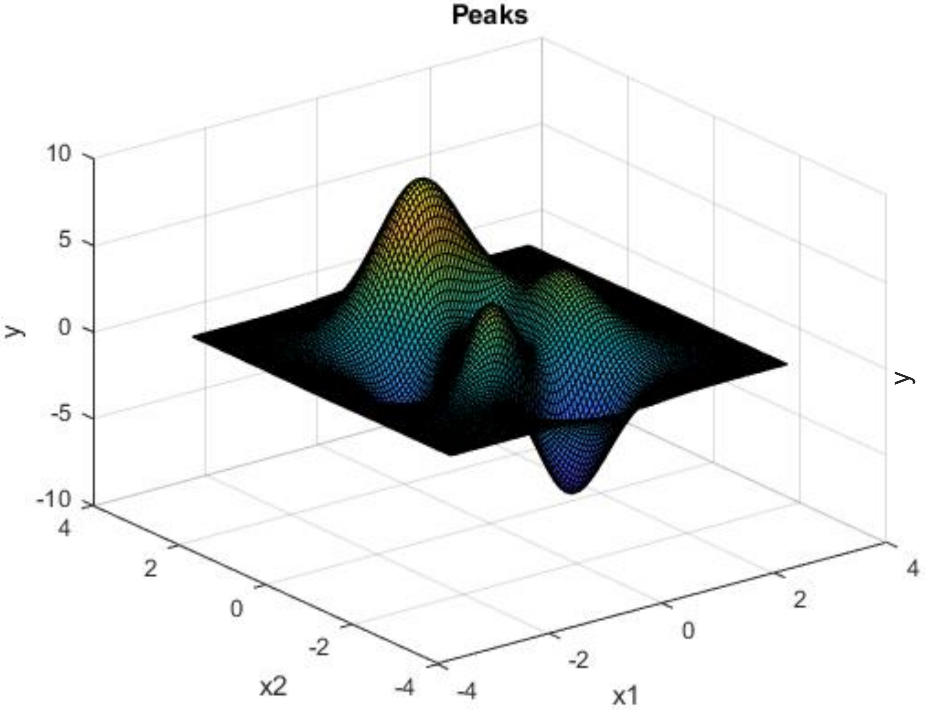
Non-operational Design Space with Data Classification

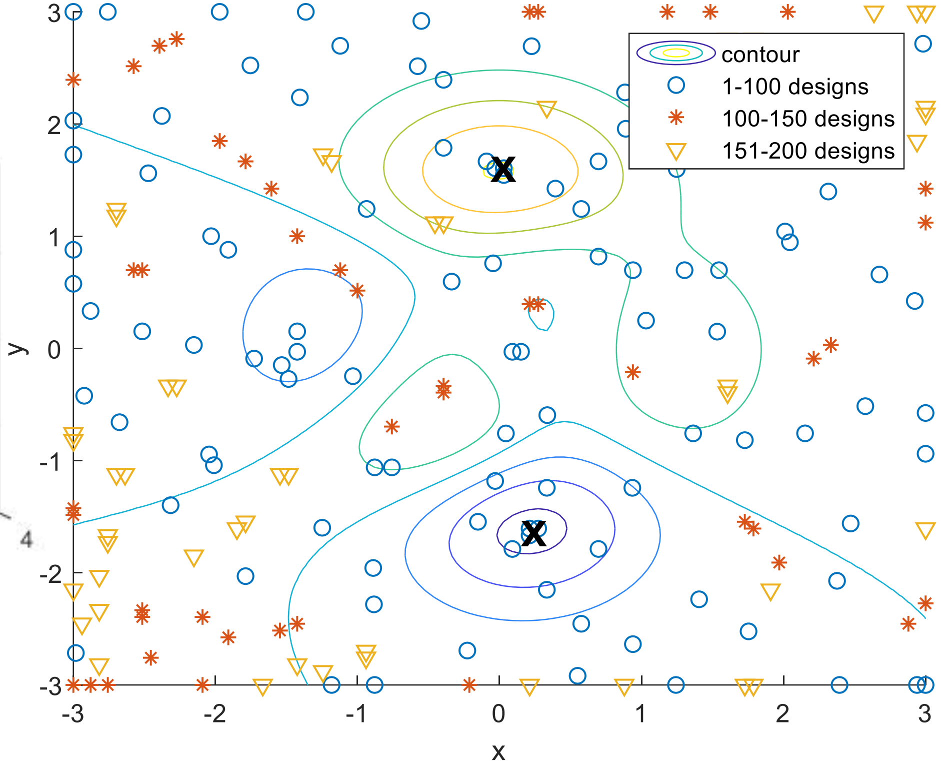


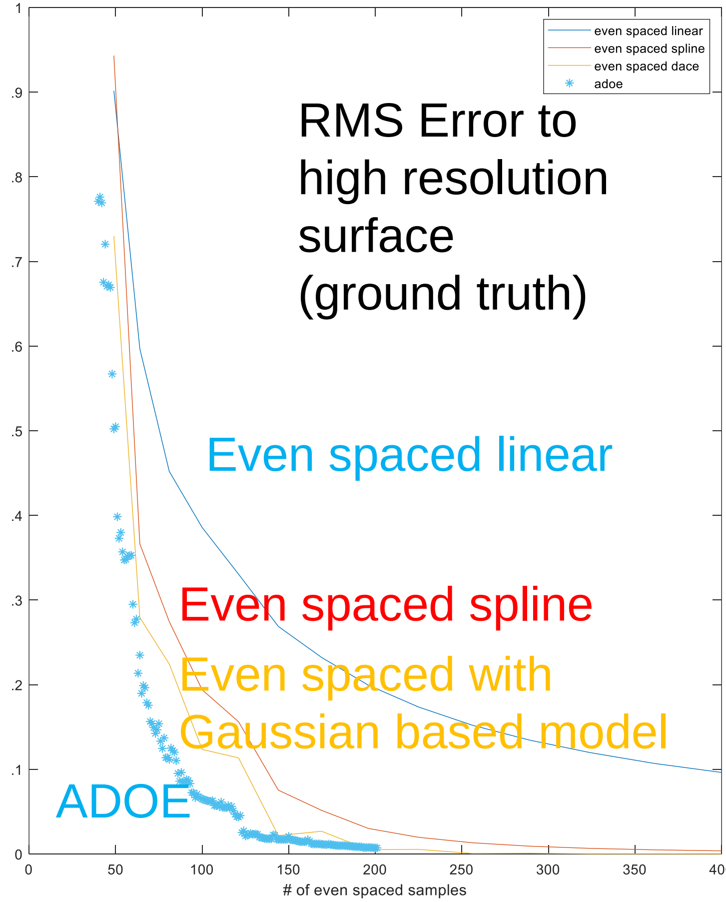
Classification helps limit evaluation space for design

- A lot of systems have non-operational design spaces
- Adaptive designs need to focus in the areas of interests and any new designs in the nonoperational spaces should be avoided to improve efficiency
- Evaluation space for the surrogate model learns the operational and nonoperational spaces with **online** data classification.
- Classification is more aggressive in the operational space to guarantee coverage.
- The same works for constraint handling

Adaptive DoE Numerical Example: Low Dimension High Nonlinearity

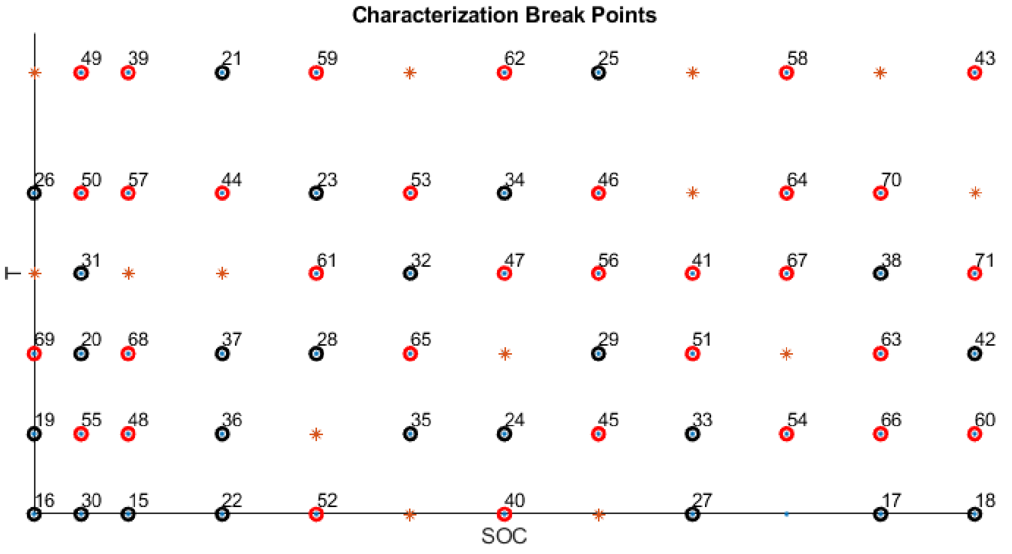


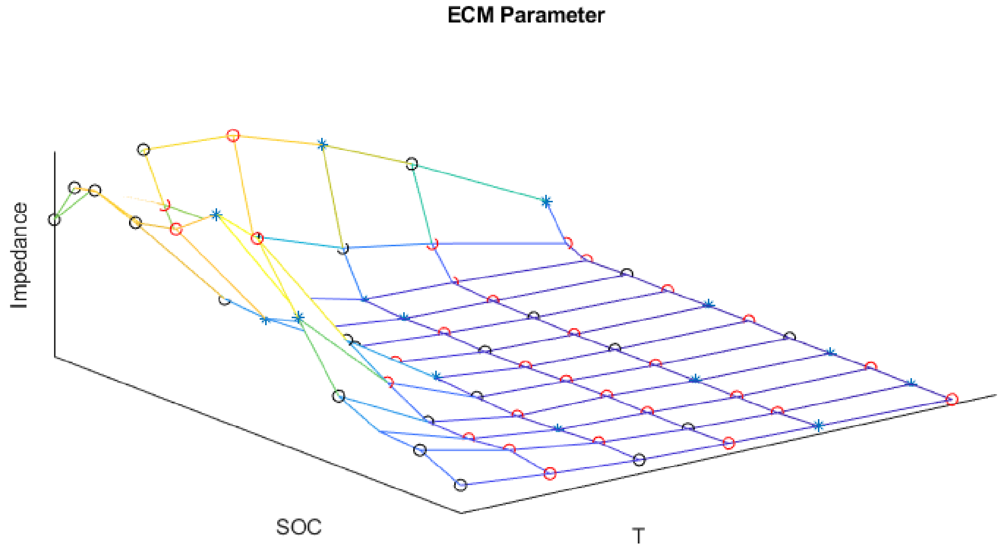


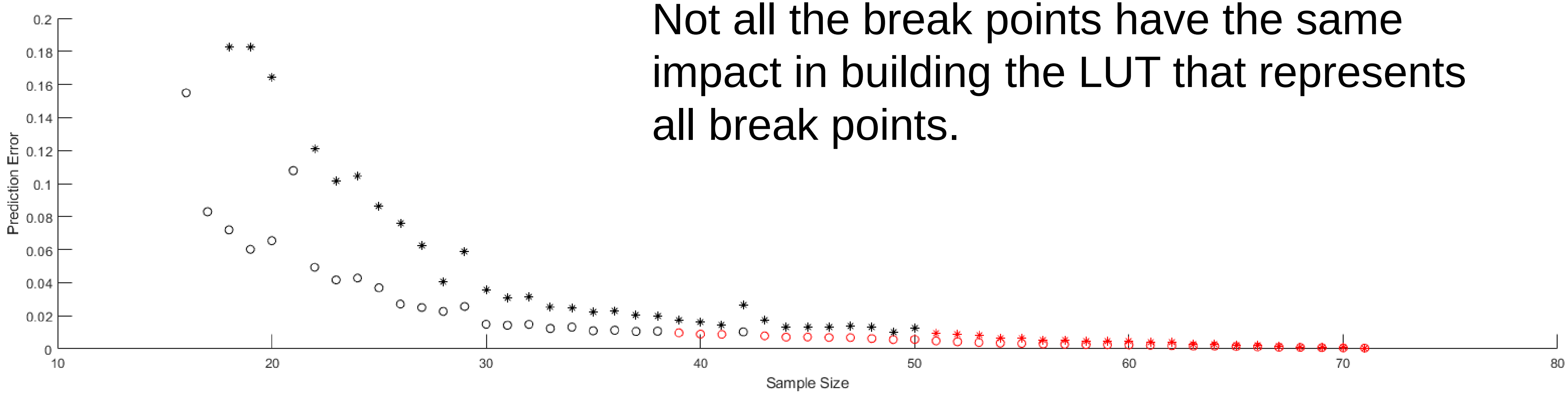


- ➢ **Efficient identification of nonlinearity and optimality**
- ➢ **Consistently better surface representations with adaptive DoE when the # of samples are limited**

Adaptive DoE Application to Reduce Battery Testing

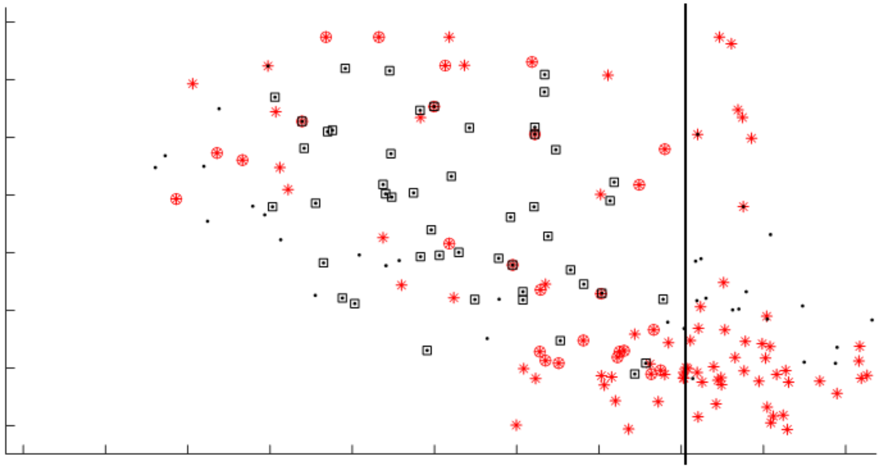


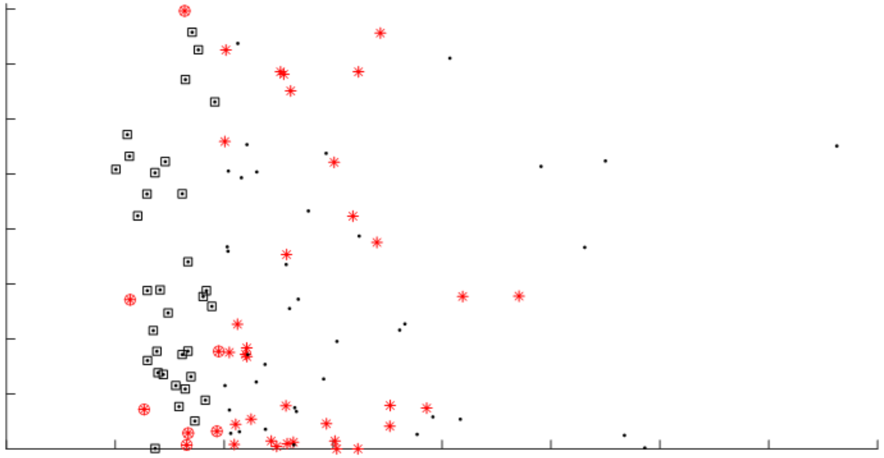


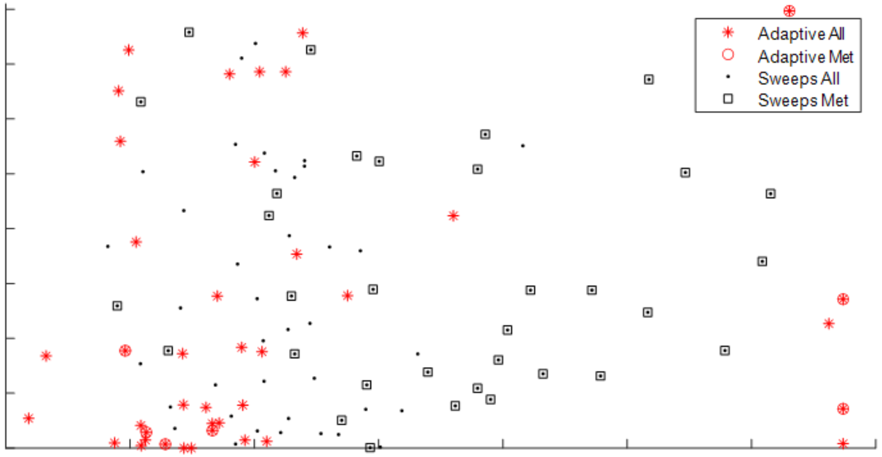


Not all the break points have the same impact in building the LUT that represents all break points.

Adaptive DOE Gasoline Application: Local DoE





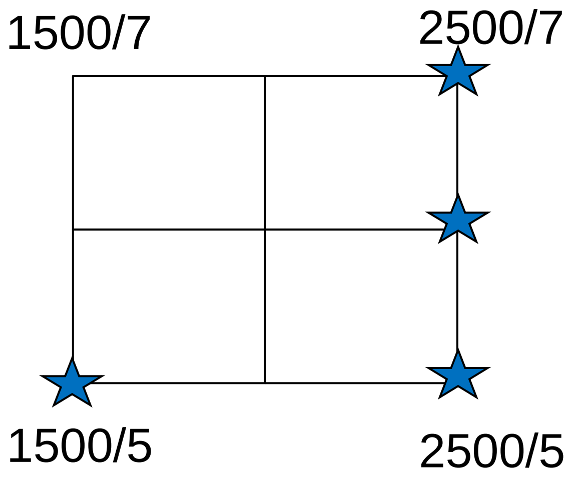


- Comparable pareto coverage in the range meeting all constraints
- Single point DoE achieved similar optimality with similar amount of tests designed based on experience, though the process took out human intervention and reduces engineering time.

Regional DoE

To further demonstrate the benefit, we explored regional DoE

- Adjacent operation points have similar and smooth characteristics.
- Building regional surface could reduce amount of data needed
- N operating points don’t need N times data needed at individual operation points



- Proof of concept
  - Look at 9 operating points (1500/5 – 2500/7) with adaptive DoE
  - Run / compare with traditional sweeps

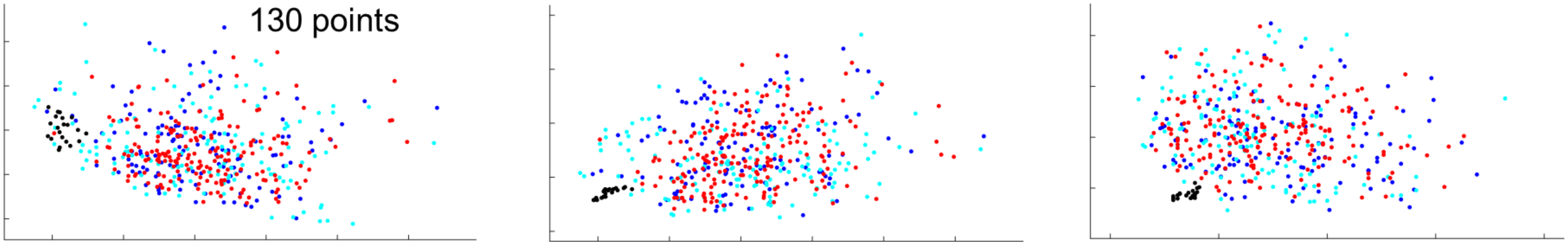
Test Performance & Efficiency

- For all individual operating points, the adaptive DoE found the same calibration as the traditional methods

	Extended Range DOE	**1500 RPM/5 bar**	**2500 RPM/5 bar**	**2500 RPM/6 bar**	**2500 RPM/7 bar**
Total # of Test Points	569	92	75	68	62

- The four conventional mapping/validation studies incorporated between 62 and 92 points, for an average of ~74 points per speed/load point with human intelligence to reduce the amount of data for subsequent test points.
- Extended range DOE consisted of 569 test points, or ~63 points per speed/load point
  - 15% reduction in number of points needed
  - Significant engineering time reduction to screen and analyze data can be achieved by adaptive DoE

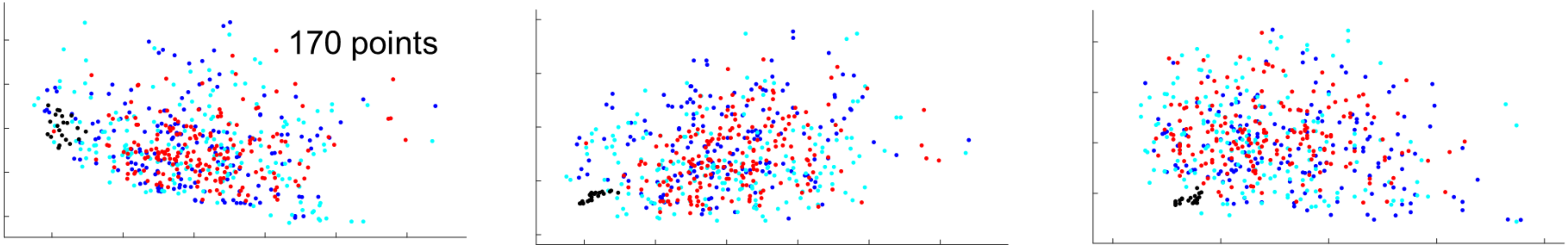
Robustness Results: Repeated Local DoE Tests



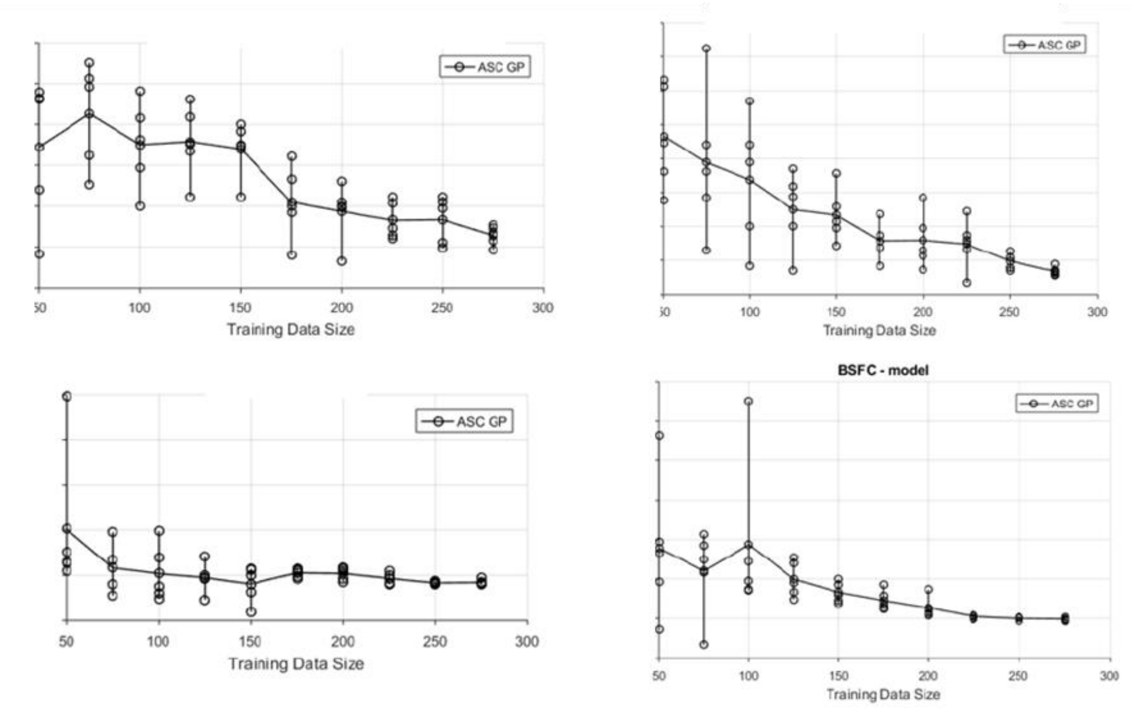
- ● Adaptive DoE 1
- ● Adaptive DoE 2
- ● Base DoE
- ● Validation Points

Adaptive 1: 100 initial designs
Adaptive 2: 70 initial designs

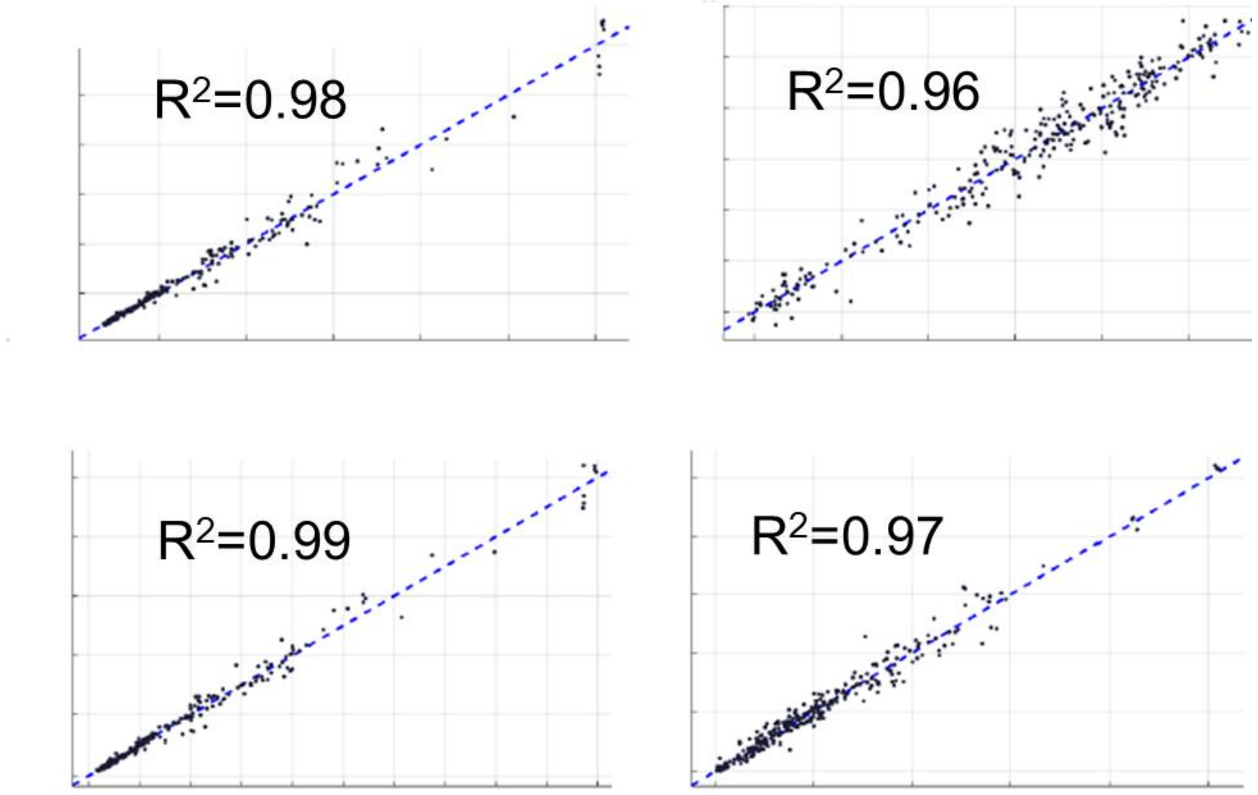
- ➢ **Repeated adaptive runs achieved similar coverage of base designs + validation**
- ➢ **Starting with smaller initial design size leads to faster convergence / less data needed (covering optimality with 170 points instead of 250 points)**



Modeling Data Evaluated by Commercial Tool



Adaptive DoE training data modeled using commercial tool to evaluate model quality for regional DoE (10 design variables)



- High R^2 convergence observed on all models
- Model errors sensitivity on training data size shows mean error plateau at approximately 250 training data size, indicating that 250-300 training data is sufficient

Conclusions

- Adaptive DoE is an efficient test procedure to incorporate design of experiments with hardware operation
  - Simultaneously provide data for both modeling and calibration
  - Benefits to optimality and efficiency
  - Free up engineering resource needed for current manual process
  - Consistent process
  - Higher benefits for high DOF and high nonlinearity problems.
  - Useful for expensive (cost and/or duration) tests, including physical testing and virtual testing
- MATLAB is the only tool used to develop the whole package

MathWorks
AUTOMOTIVE CONFERENCE 2023
North America

Thank you

MathWorks®