

Applications and Extensions of Response Surface Methodology at NASA

Peter A. Parker, Ph.D., P.E.

National Aeronautics and Space Administration

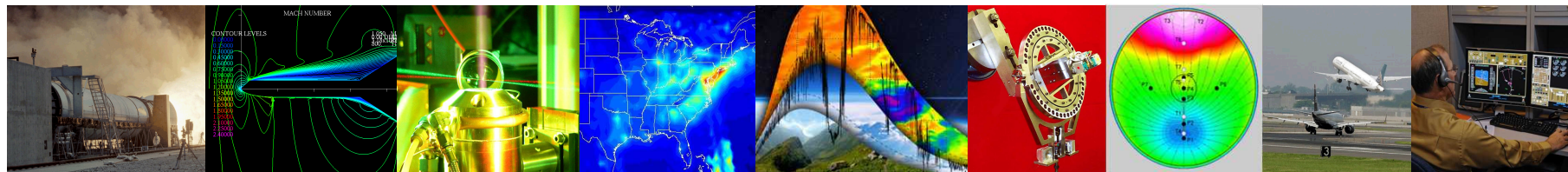
Langley Research Center

Hampton, Virginia, USA

February 26, 2019

NIST Statistical Engineering Division Seminar

Gaithersburg, MD



NASA Langley Research Center

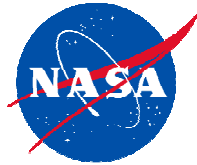


- **One of 10 NASA Centers, located in Southeast Virginia**
- **Established in 1917, first civilian aeronautics laboratory in United States**
- **Initial home of the first astronauts (Mercury 7)**

Today, we focus our research on technical challenges in:

- **Space Exploration**
- **Revolutionizing Aeronautics and Air Transportation**
- **Global Climate Change**





Overview of Presentation

- 1. Background and brief survey of statistical applications at NASA**
- 2. Computational RSM with higher-order models in acoustic classification of environmental noise sources**
- 3. Strategies for near replicates in response surface methodology in measurement system characterization**

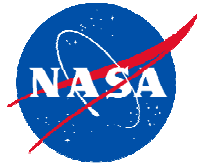


Background

- **Response surface methodology was introduced in 1951, wide spread industry usage by the 1980's**
 - **However, RSM application in aerospace research and development is relatively new (last 20 years) and still growing - Why?**
- **Traditional academic training of engineers lacked relevant, applicable, and implementable statistical courses**
 - **Most engineers were unaware of industrial, applied statistics**
- **Engineers were (are) generally trained to "fear" uncertainty**
 - **Uncertainty is a problem, a nuisance to be avoided**
 - **"Never replicate an experiment, because it causes error."**
- **Over last 15 years, there are better educational resources, accessible software, and an increasing appreciation for the role of statistical methods in aerospace research**

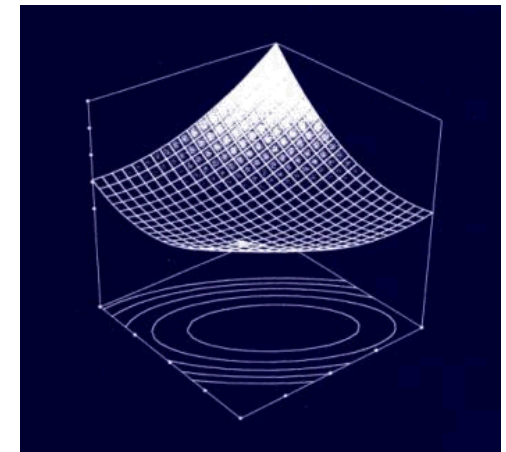
Statistical methods enable efficient, insightful, data-driven decisions in the presence of uncertainty

Some Predominant Statistical Methods



- **Design of Experiments (Physical and Computational)**
- **Response Surface Methodology (RSM) (Physical and Computational)**
- **Modeling, Regression, Nonparametric**
- **Stochastic Simulation**
- **Statistical Process Monitoring, Statistical Quality Control**
- **Uncertainty Analysis and Propagation**

**Methods for rigorously understanding
factor-response relationships**



A Few “Radical” Statistical Concepts

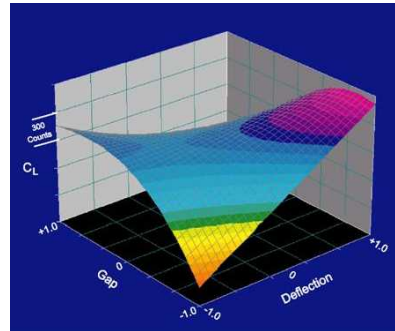
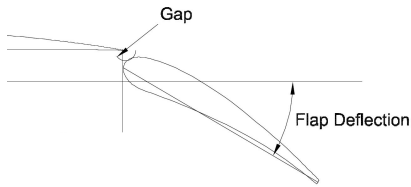


- **Experiment's are designed**
 - Starts with precise questions, risks, and consequences
 - Analysis plan is developed and experiment sufficiency is evaluated before the data are collected
- **Factorial experiments are effective**
 - Changing more than one factor at a time, simultaneously maximizes information acquisition efficiency
 - Interactions provide richer insights
- **Modeling with quantified uncertainty is defensible**
 - Estimating and isolating random variability from modeling bias supports decisions
 - Overcomes weakness subjective assessment of subtle differences in factor-response relationships

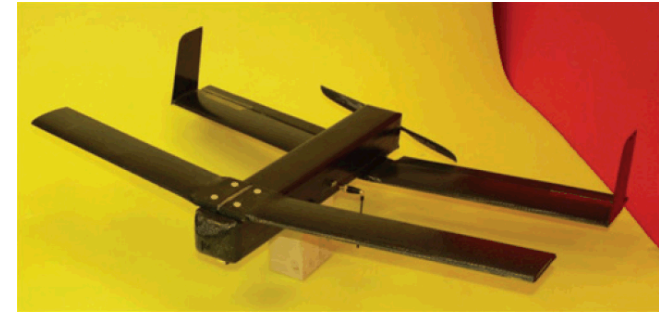
Techniques to efficiently and sequentially learn



Aerosciences Wind Tunnel Testing



DeLoach, et al. (2001), "Practical Applications of Response Surface Methods in the NTF," AIAA 2001-0167



English, et al. (2012), "An Efficient Split-Plot Approach for Modeling Nonlinear Aerodynamic Effects," *Quality Engineering*, 24: 522-530

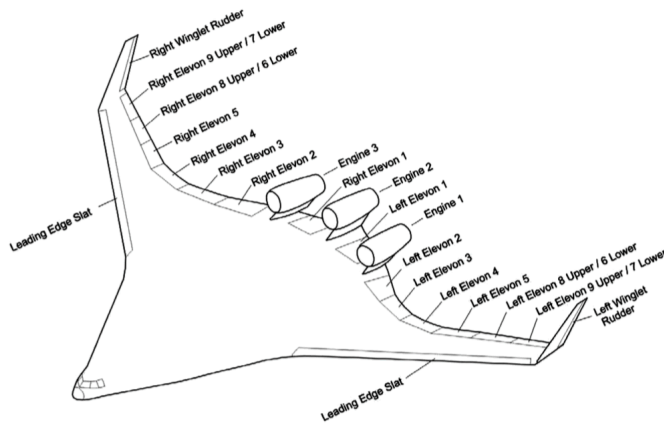


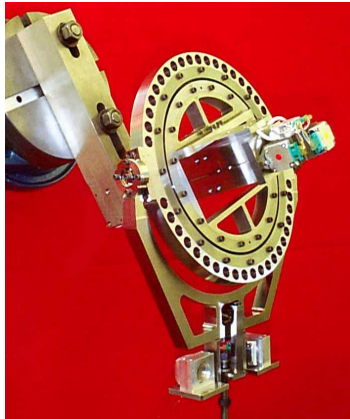
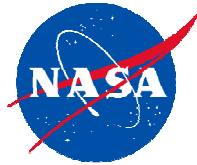
Fig. 2 BWB configuration control surfaces.

Landman, et al. (2007), "Response Surface Methods for Efficient Complex Aircraft Configuration Aerodynamic Characterization," *Journal of Aircraft*, 44, (4)



Overmeyer, et al. (2015), "Case Studies for the Statistical Design of Experiments Applied to Powered Rotor Wind Tunnel Tests," AIAA 2015-2713

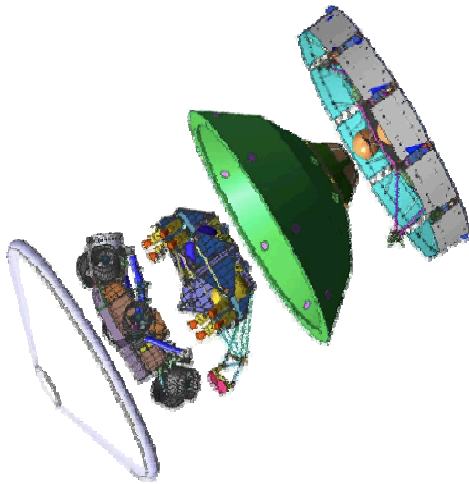
Measurement System Characterization



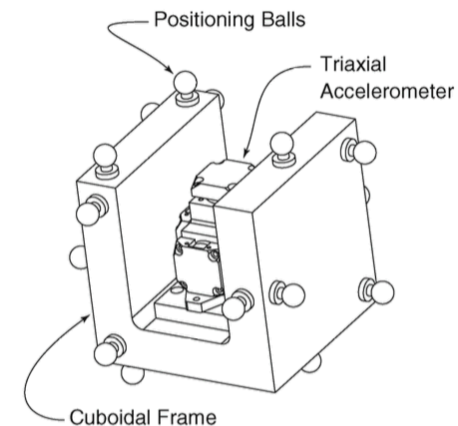
Parker, et al. (2001), "A Single-Vector Force Calibration Method Featuring the Modern Design of Experiments," AIAA 2001-0170



Gentz, Parker, Rhew, Roberts (2006), "Solid Rocket Motor Characterization of Ares I." NASA TM

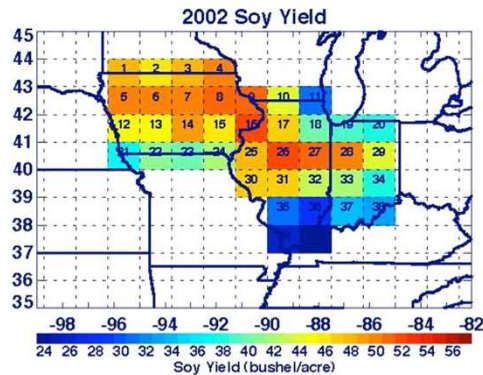


Parker and Commo (2012), "Statistical Engineering Perspective on Planetary Entry, Descent, and Landing," *Quality Engineering*, 24, 2



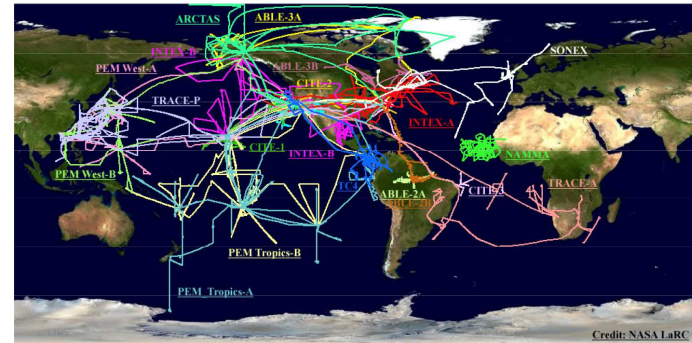
Parker and Finley (2007), "Advancements in Aircraft Model Force and Attitude Instrumentation by Integrating Statistical Methods," *Journal of Aircraft*, 44, (2)

Atmospheric Sciences

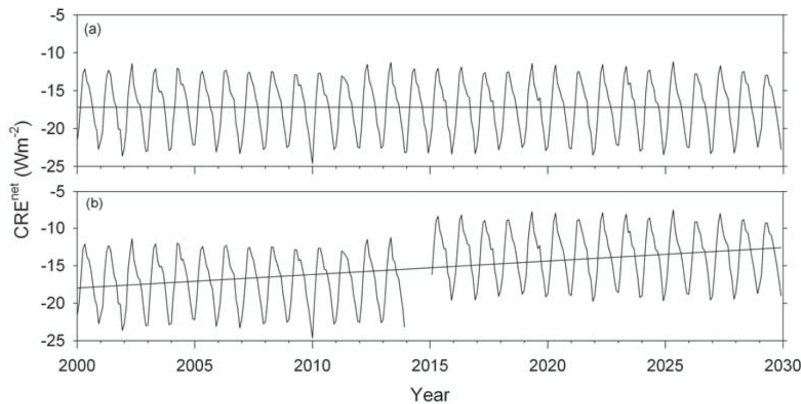


Fishman, et al. (2006), "An Investigation of Widespread Ozone Damage to the Soybean Crop in the Upper Midwest Determined from Ground-Based and Satellite Measurements," *Atmospheric Environment*, 44, pp. 2248-2256.

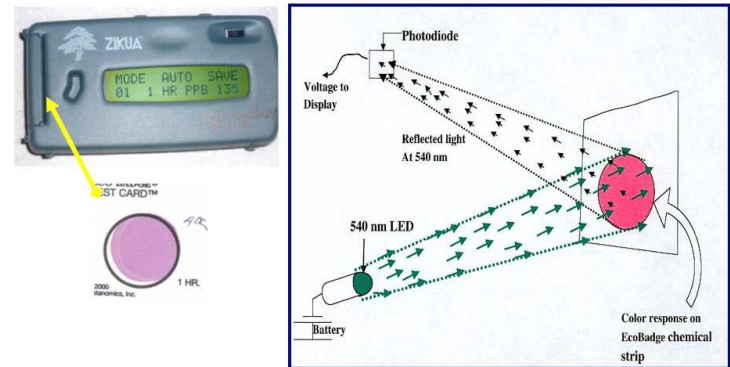
NASA Tropospheric Airborne Campaigns (1983 – 2008)



Kleb, et al. (2008), "Creating Airborne ESDR Products for Global and Regional Model Assessment," *American Geophysical Union Annual Conference*.

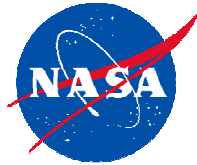


Loeb, et al. (2009), "Impact of Data Gaps on Satellite Broadband Radiation Records," *Journal of Geophysical Research*, 114.



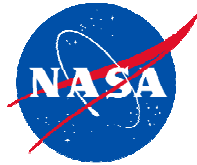
Pippin, et al. (2007), "Improvements to the Passive Ozone Measurement System Used by GLOBE Schools," *American Geophysical Union Annual Conference*.

Motivations for RSM Extensions



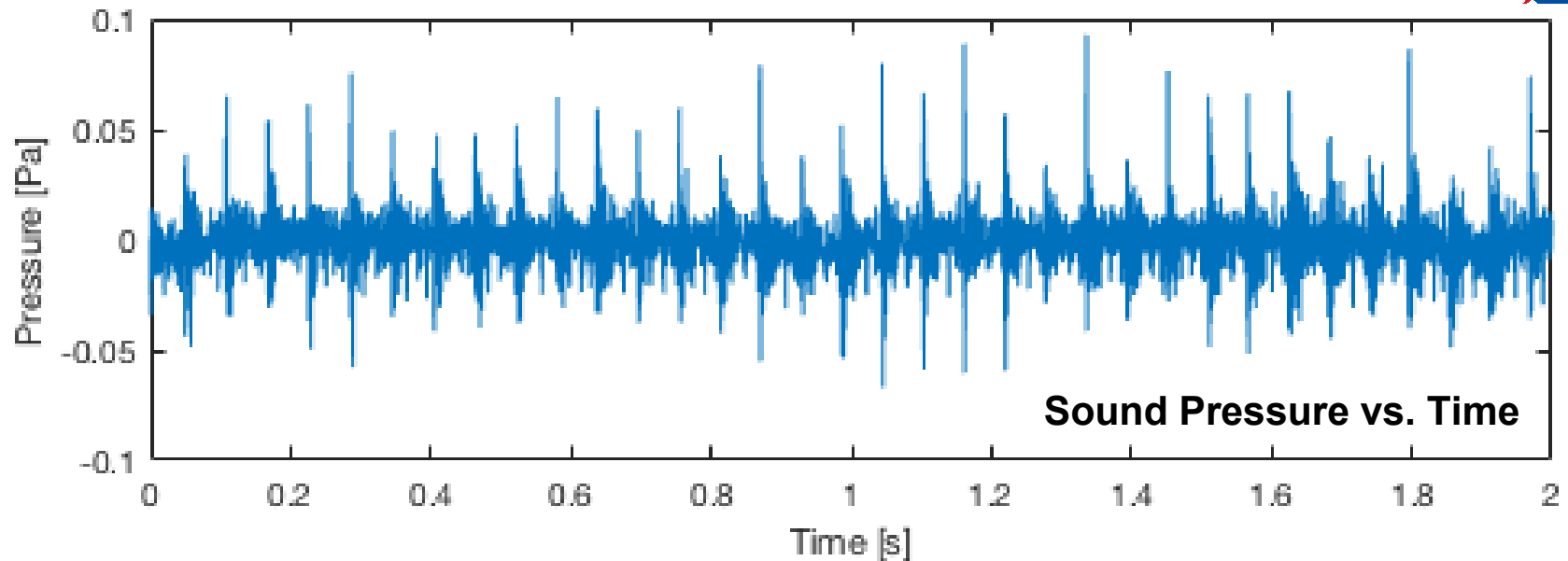
- **RSM was developed for industrial product and process optimization**
 - **Objective to identify “optimal” factor settings to maximize yield, minimize waste – ultimately, to maximize corporate profits and/or market share**
- **At NASA, RSM is used to accelerate learning, maximize knowledge, and inform data-driven decisions in aerospace research and development**
 - **Fundamental approach and objectives are similar to classical RSM, however extension is often required for the aerospace class of applications**
- **Two specific examples of RSM adaptations and extensions at NASA**
 - **Computational experiments exploring higher-order models**
 - **Estimation of pure-error with near replicates**

Acoustic Classification of Environmental Noise Signatures

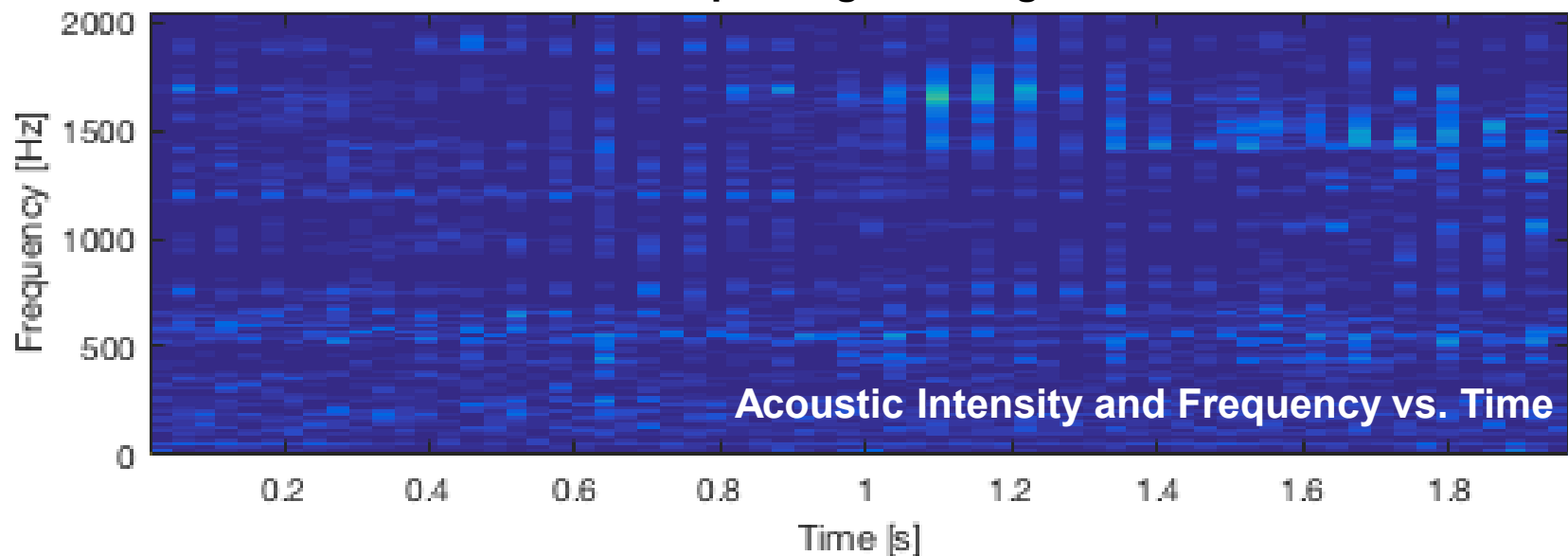


Goal: Develop an Environmental Noise Classifier System

Acoustic Signals to Spectrogram Image

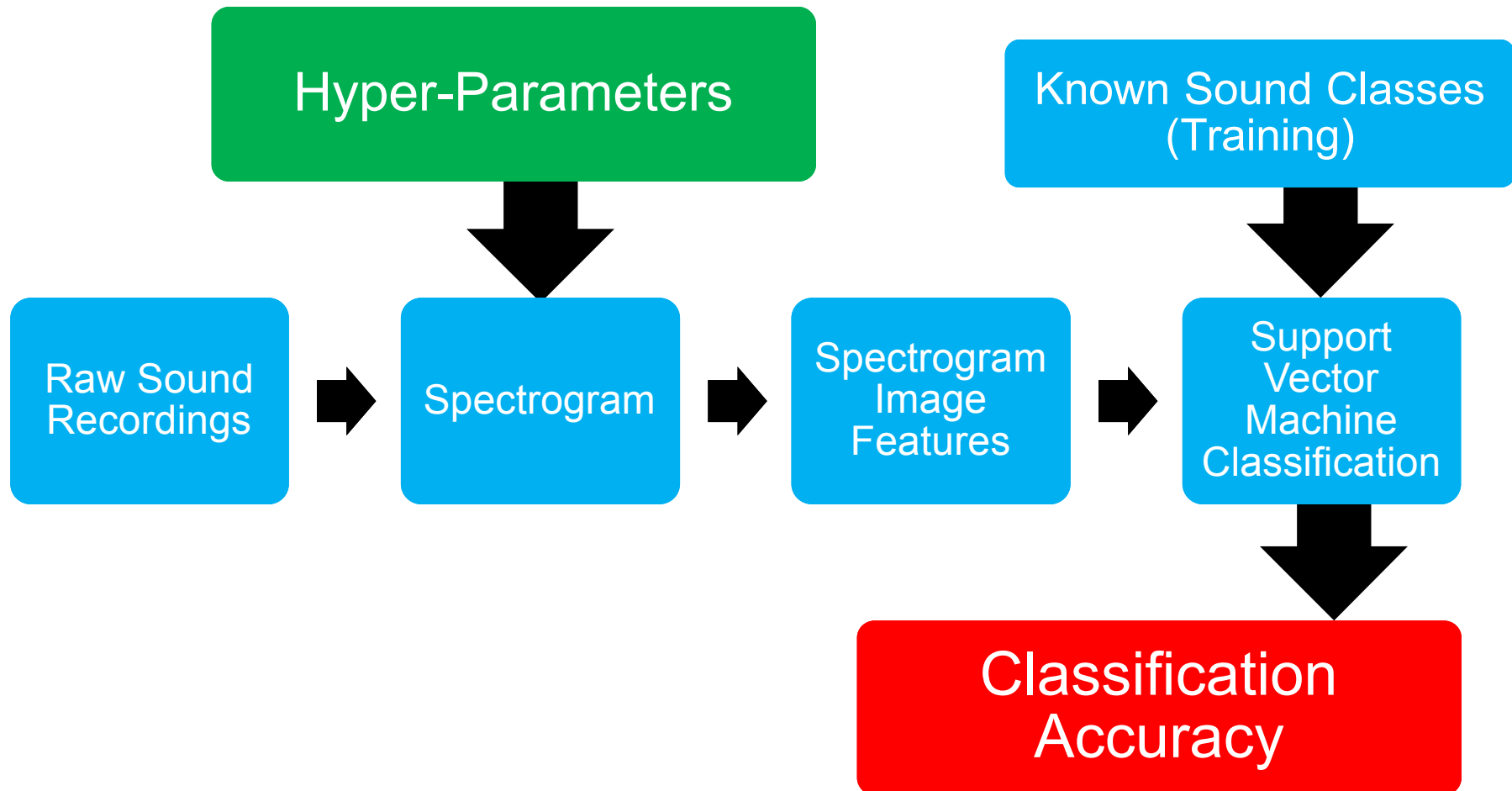


Spectrogram Image





Classification System



Identify Spectrogram Hyper-Parameters that improve the Classification Accuracy and reduce Computational costs

Experimental Factors and Responses



Factors

Hyper-parameters

x1: # FFT Points

x2: # Time Bins

x3: % Window Overlap

Previous attempts explored:

8 levels of x1 & x2

100 levels of x3

= 6400 combinations

Responses

Classification Accuracy

y1: Air conditioner

y2: Children playing

y3: Dog barking

y4: Drilling

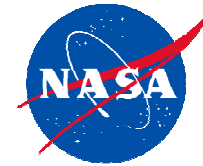
y5: Engine idling

y6: Jackhammer

y7: Street music

y8: Average over all

z1: Computational Time

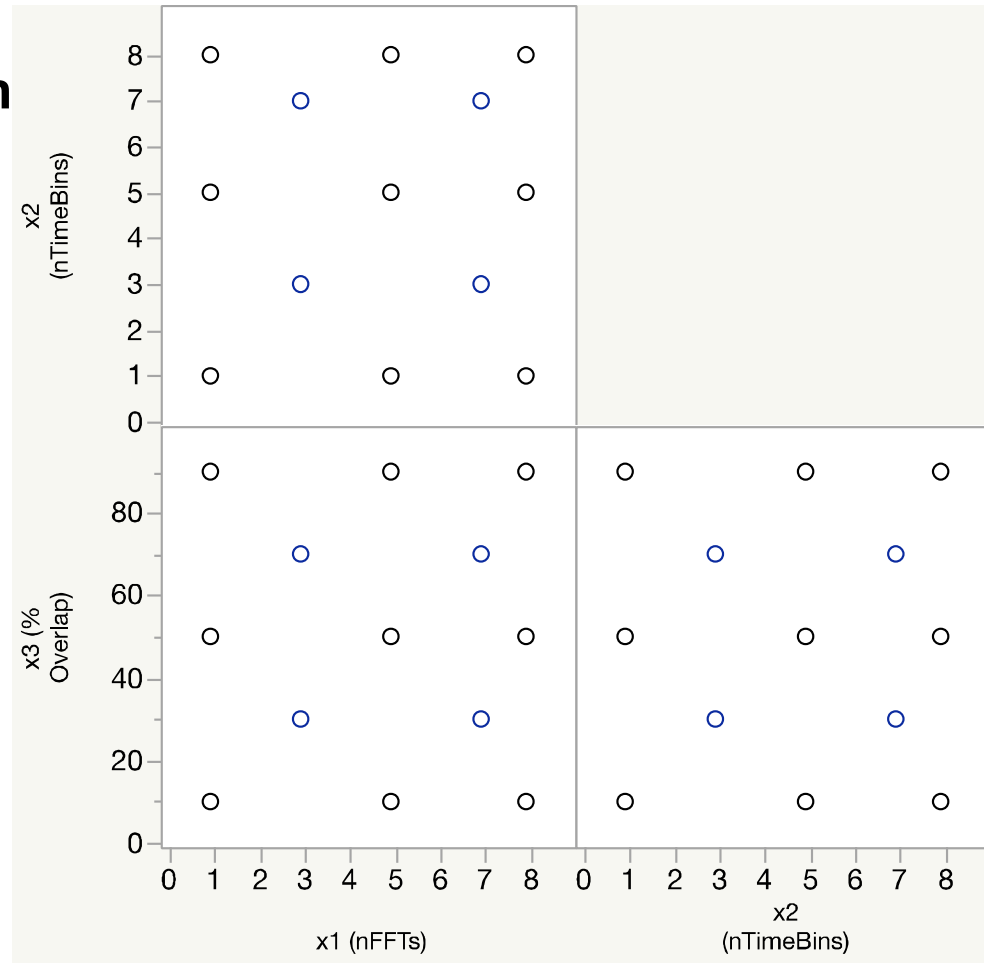


Experimental Design

- **Superimposed Face-Centered Central Composite Design with Inscribed Factorial**

25 design points

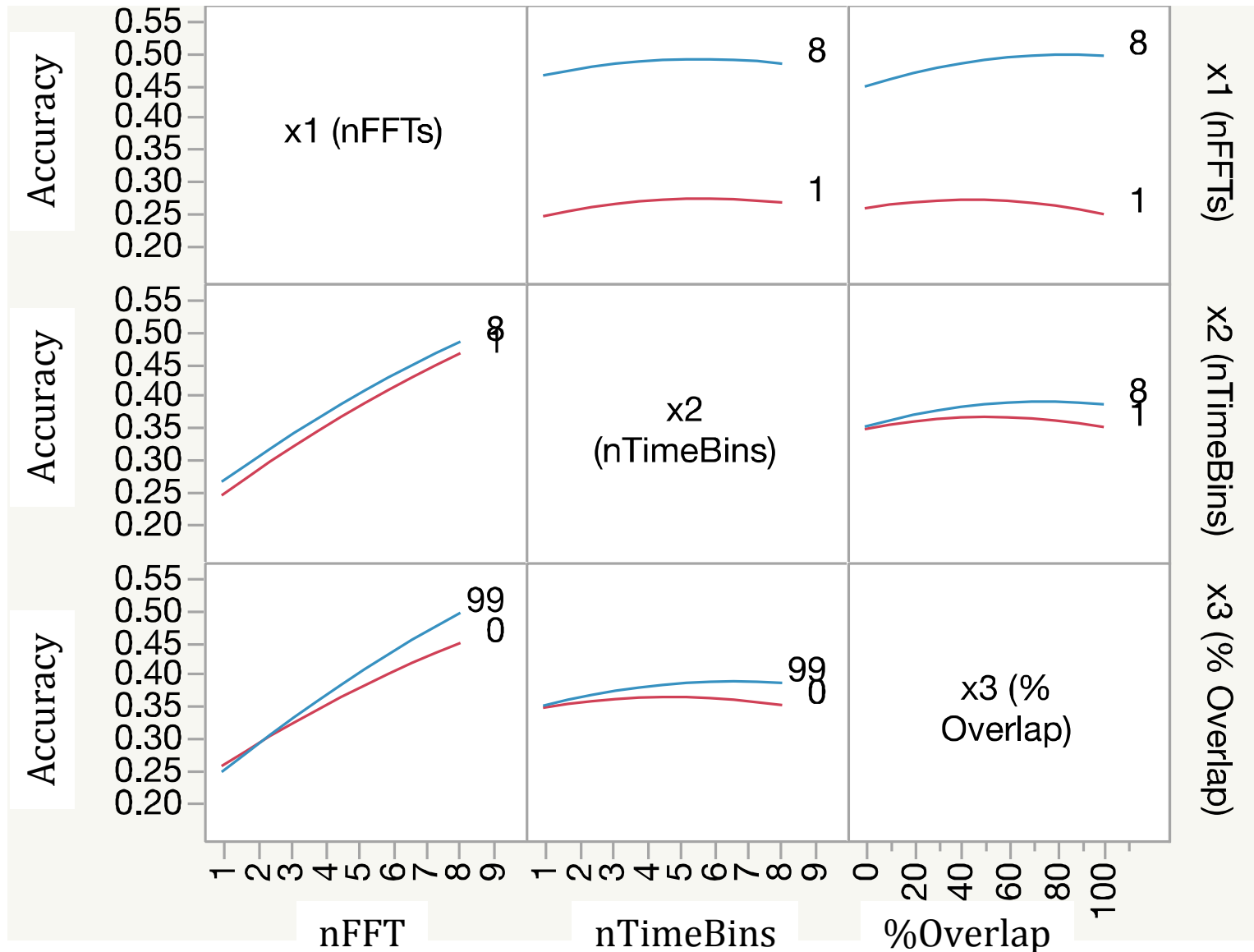
- 14 face-centered axial and factorial points
- 8 inscribed factorial points
- 3 center points to assess variability in 5-fold cross validation
- 5 unique levels of each factor
- Supports second-order model, with mixed higher-order terms



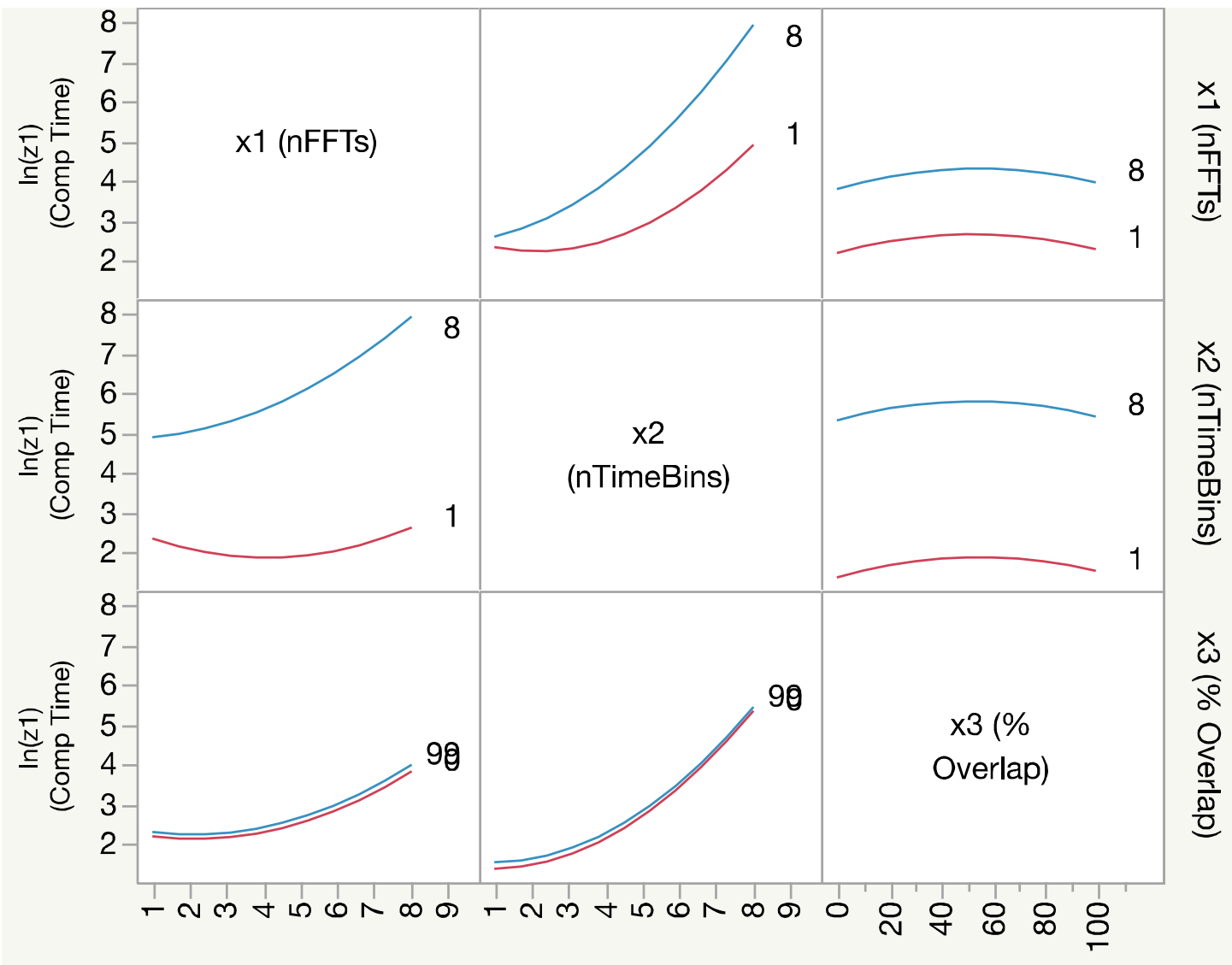
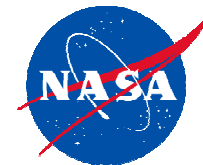
$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^k \beta_{ij} x_i x_j$$

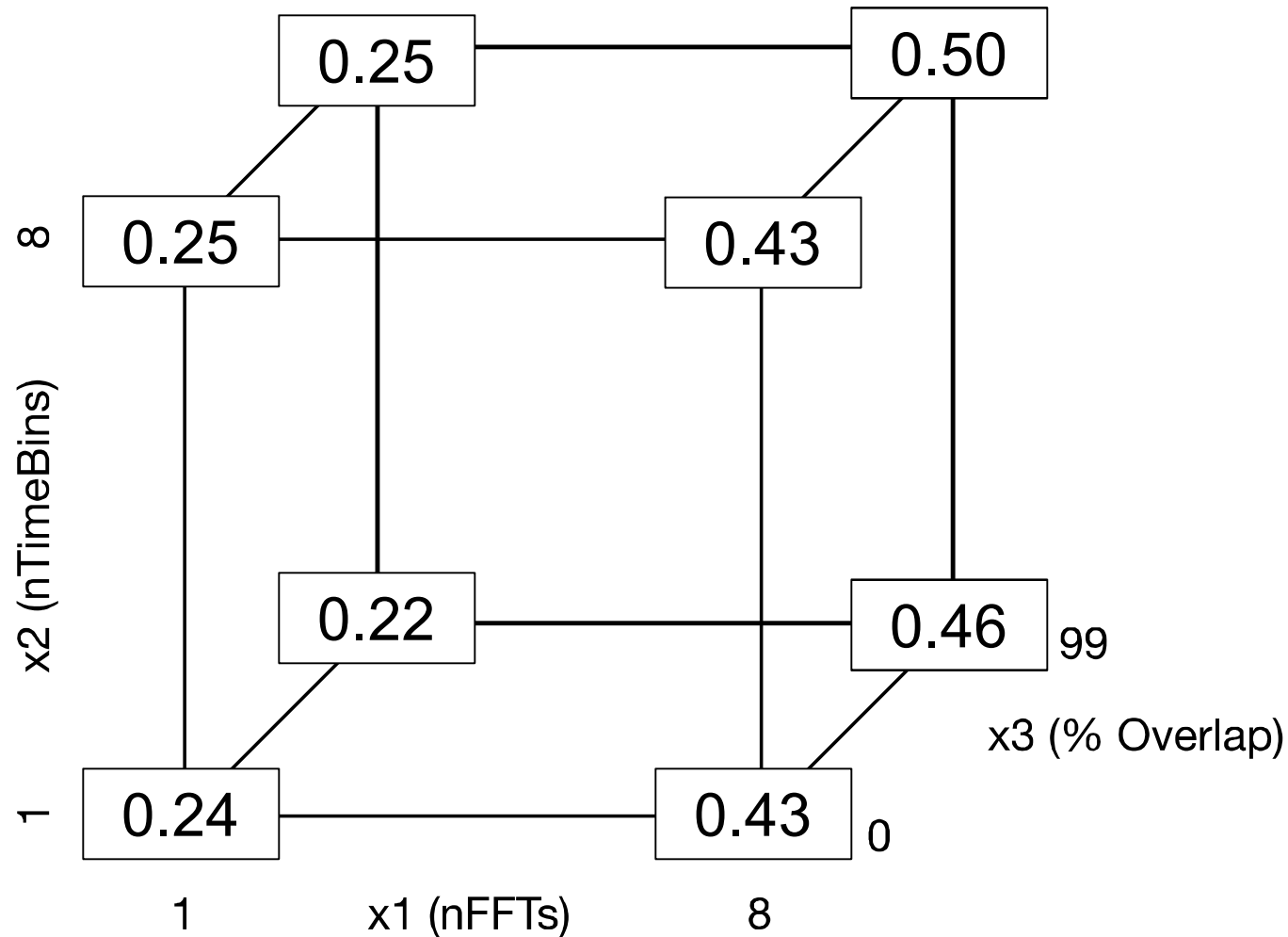
Interaction Plot

Average Classification Accuracy (y8)

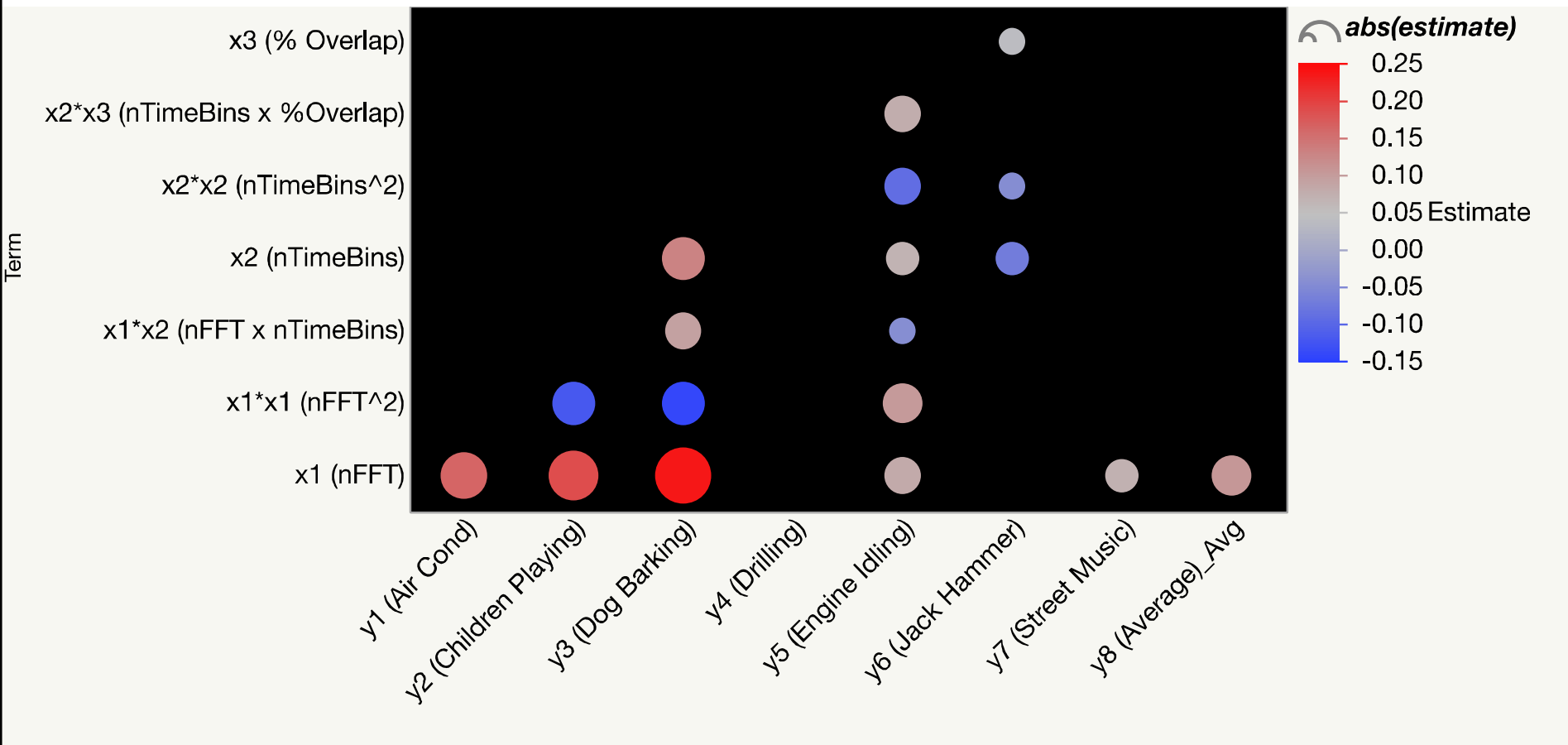
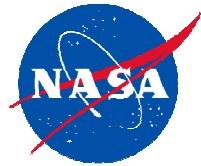


Computational Cost Model, $\ln(z1)$





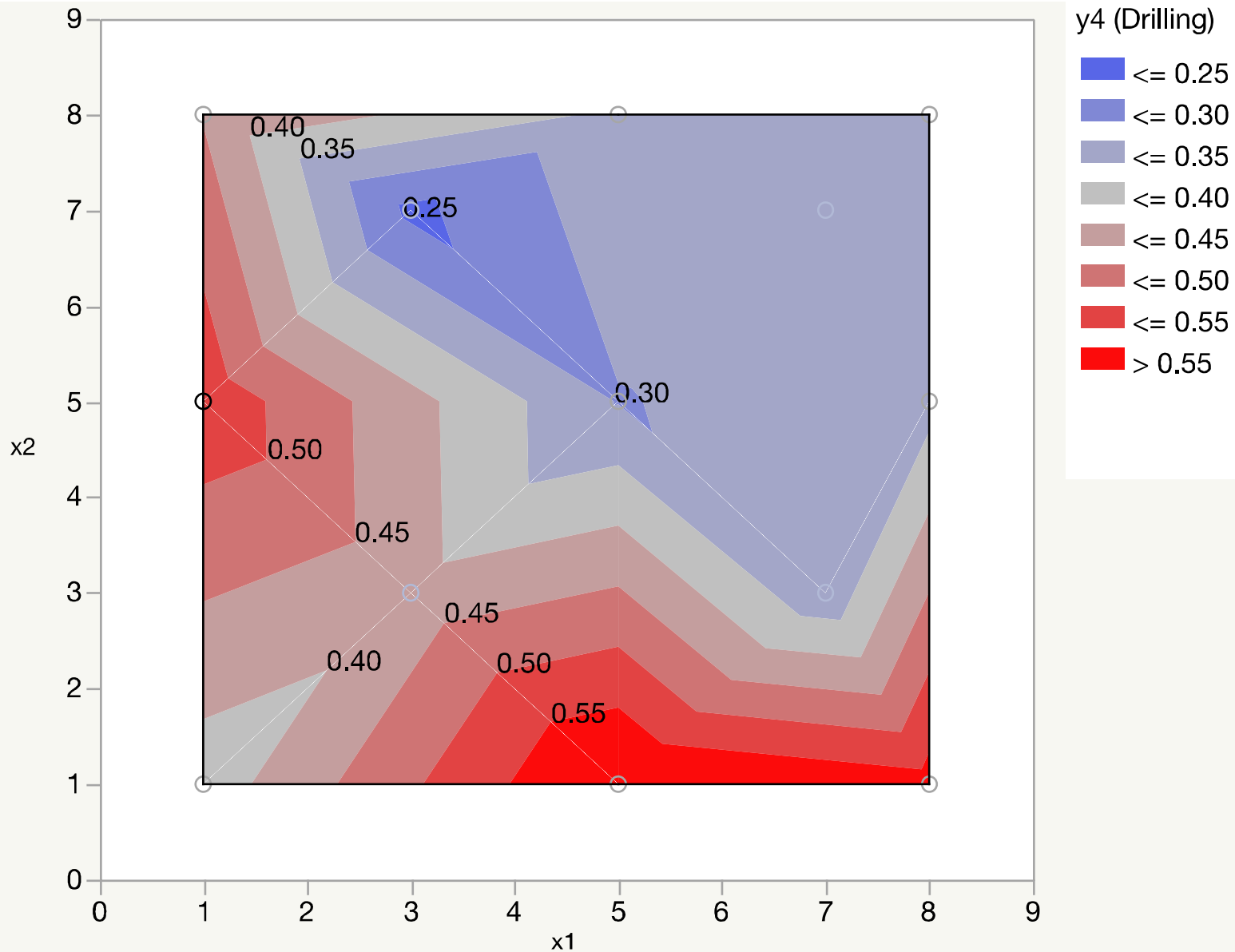
Models for Each Noise Classification Exploring their Magnitude and Polarity



Dot size represents Relative magnitude of effect

Color indicates Effect Polarity – Red directly proportional, Blue inversely proportional

Contour Map of Raw Data of y4 (Drilling)



Augmented 2nd Order Model with Select 3rd and 4th Order Terms

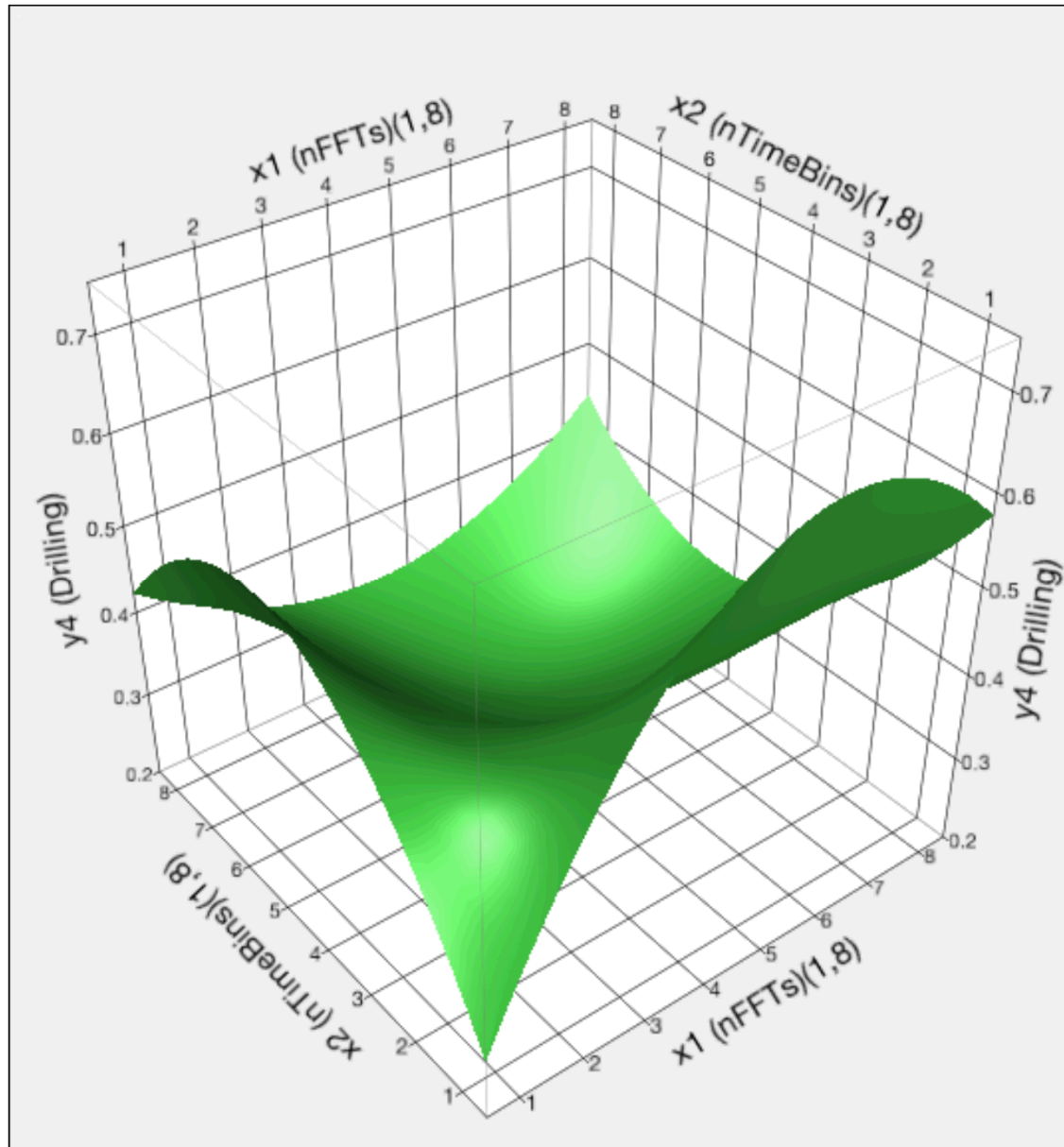


Term

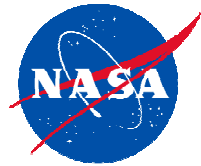
- * $x_1(n\text{FFTs}) * x_2(n\text{TimeBins}) * x_2(n\text{TimeBins})$
- $x_1(n\text{FFTs}) * x_2(n\text{TimeBins})$
- $x_2(n\text{TimeBins})(1,8)$
- $x_1(n\text{FFTs})(1,8)$
- * $x_1(n\text{FFTs}) * x_3(\% \text{ Overlap}) * x_3(\% \text{ Overlap})$
- $x_3(\% \text{ Overlap}) * x_3(\% \text{ Overlap})$
- $x_1(n\text{FFTs}) * x_1(n\text{FFTs})$
- * $x_1(n\text{FFTs}) * x_1(n\text{FFTs}) * x_2(n\text{TimeBins}) * x_2(n\text{TimeBins})$
- * $x_1(n\text{FFTs}) * x_1(n\text{FFTs}) * x_2(n\text{TimeBins})$
- $x_3(\% \text{ Overlap})(0,99)$
- $x_2(n\text{TimeBins}) * x_2(n\text{TimeBins})$

* Terms above 2nd Order

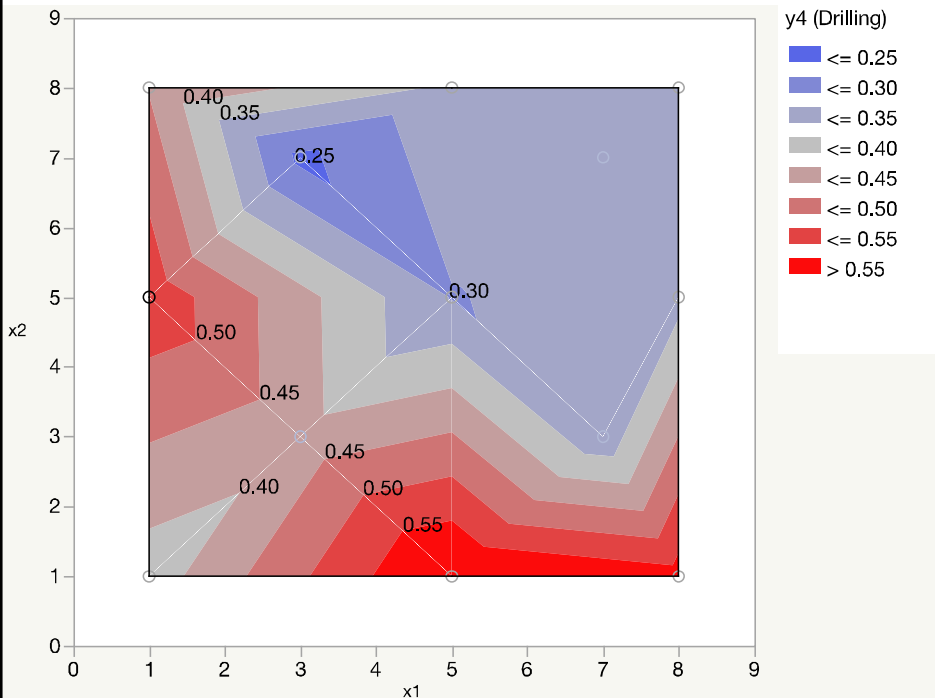
Drilling (y4) 4th Order Response Surface



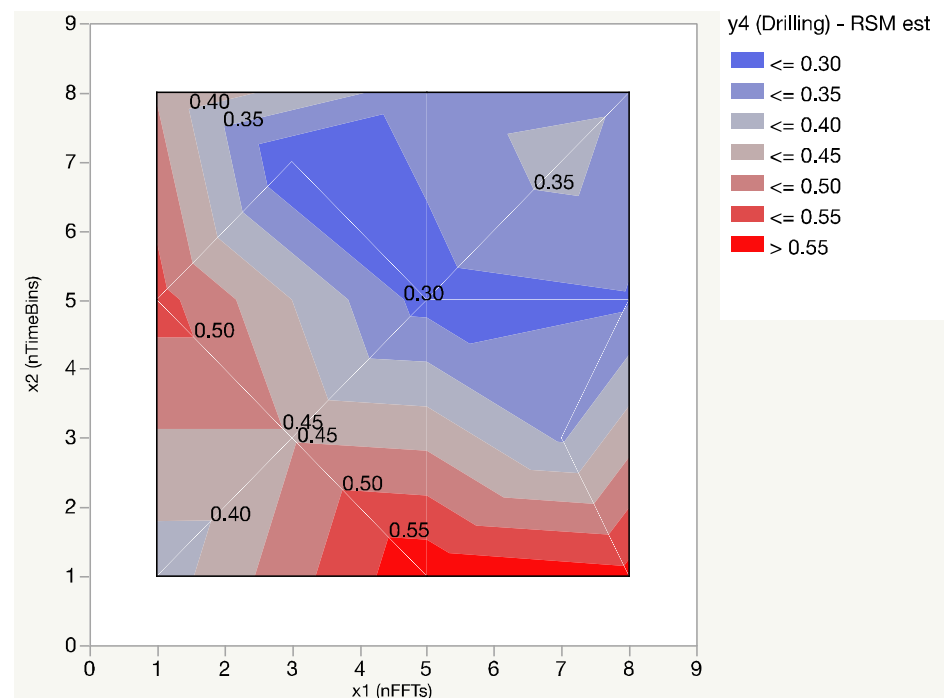
Comparing Raw Data and RSM Model



All 25 RSM Points



Predicted 4th Order RSM Model





Acoustic Classification Results

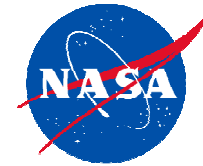
- **RSM developed simple, parsimonious models that:**
 - Adequately explore the hyper-parameter space
 - Identify important factors and relationships
- **RSM proved to be an effective and efficient sampling method**
 - 25 points significantly less than 6400 exhaustive sampling
 - Even greater potential benefits in higher-dimensional space
- **Low order polynomial models were effective for the overall average and most of the sound classes, but not for all**

Comparing Computational Time

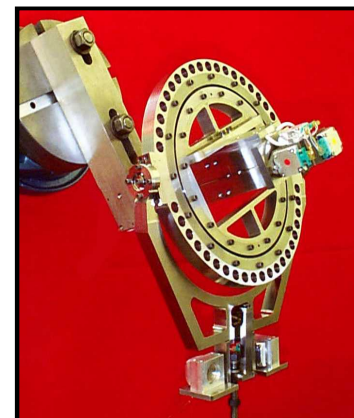
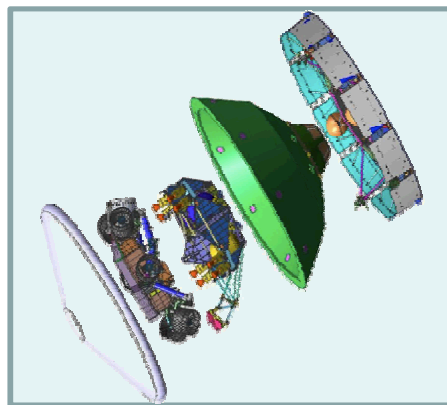
RSM: 25 design points, ~45 minutes

Other optimization methods: ~2000 samples, ~2.5 days

Exhaustive search: 6400 samples, ~8 days



Strategies to Estimate Pure-Error with Near Replicates in RSM





Near Replicates in RSM

RSM pure-error estimates are based on exact replicates of factor settings

- Powerful and highly desirable in practice
- Estimate of experimental variability, model-independent
- Useful for lack-of-fit testing, and higher-order model justification
- Easy to explain, approachable to non-statisticians

However, in practice, precise **genuine replicates are often not achieved**

- Systems may lack control precision to set the factor levels identically
- At NASA, we can often measure the factor levels of intended replicates very precisely, thereby resulting in “near replicates”
- High-precision measurement systems with accurate set-point measurements, but not precisely or efficiently controlled set-points

Some Approaches in the Literature



- Montgomery, Martin, and Peck (1980), "**Interior Analysis of the Observations in Multiple Linear Regression**," *Journal of Quality Technology*, 12(3), pg. 165-173.
- Joglekar, Schuenemeyer, and LaRicca (1989), "**Lack-of-Fit Testing When Replicates are not Available**," *The American Statistician*, 43(3), pg. 135-143.
- Su and Yang (2006), "**A Note on Lack-of-Fit Tests for Linear Models Without Replication**," *Journal of the American Statistical Association*, 101, pg. 205-210
- Christensen (1989) "**Lack-of-Fit Test Based on Near or Exact Replicates**," *The Annals of Statistics*, 17, pg. 673-683.
- Christensen and Sun (2010) "**Alternative Goodness-of-Fit Tests for Linear Models**," *Journal of the American Statistical Association*, 105, pg. 291-301.

**None are completely satisfactory solutions,
nor are they widely known or used in RSM**

Two Basic Strategies for Near Replicates



Naïve Approach – Analyze as Exact Replicates

Could lead to overestimation of pure-error if propagation of error is not appreciated and evaluated.

Compensate Near Replicates

Practical and robust, with appropriate practitioner guidance.

We propose a methodology to correct the near replicates in a manner that allows pure-error to be estimated and evaluate the performance of the proposed methodology.

Methodology Development Overview

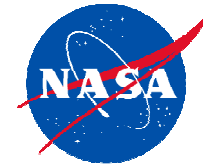


- Introduce methodology to convert **Near** to **Estimated** replicates
 - Intuitive description
 - Generalized, computational approach
- **Simulation Study** was performed to evaluate performance and robustness
 - Models Considered
 - Single-factor, First-order model without Lack-of-Fit (LoF)
 - Single-factor, First-order model with Lack-of-Fit
 - Three-factor, with Two-Factor Interactions, without Lack-of-Fit
 - Three-factor, with Two-Factor Interactions, with Lack-of-Fit
 - Varying error, lack-of-fit structure, and nearness of replicates
 - Replication of center points versus other locations in design space
 - Varying replicate degrees-of-freedom
- Simulation study explores polynomial models of varying complexity and varying degrees of LoF to understand both benefits and potential weaknesses of the methodology.

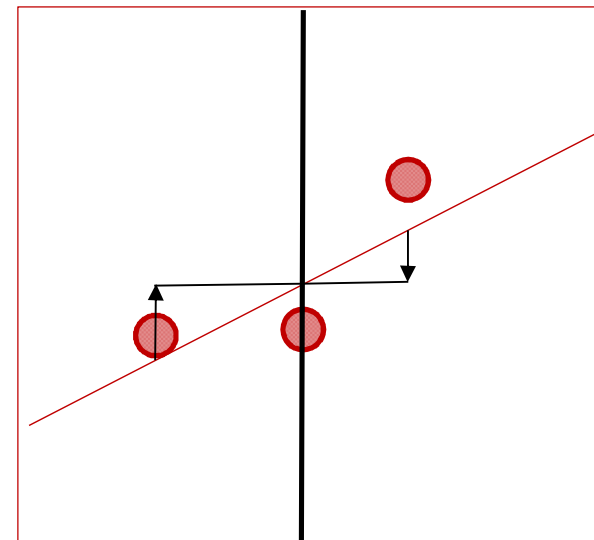
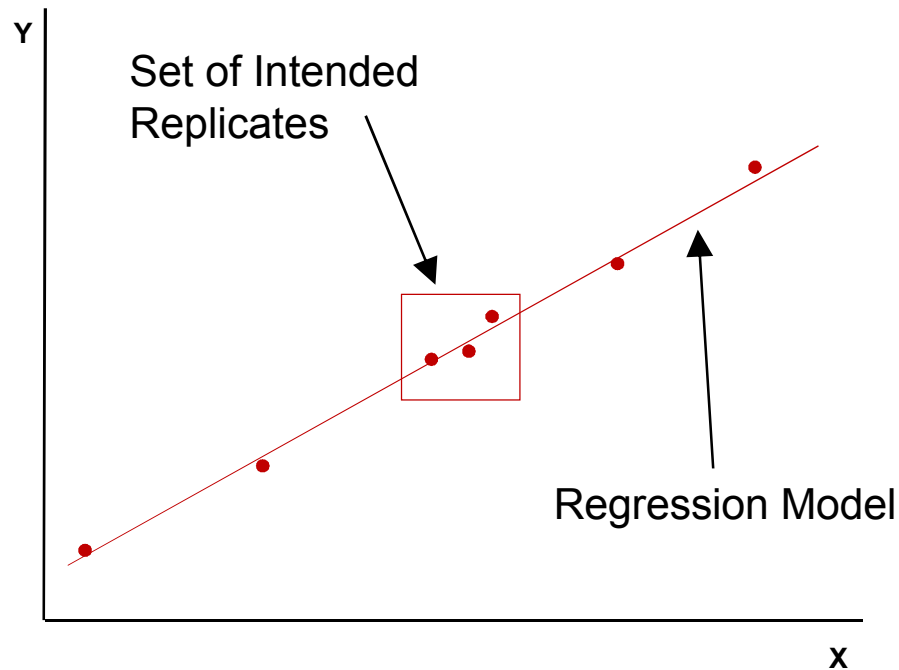


Process – Intuitive Description

- 1. Fit regression model, as in normal RSM practice**
- 2. Identify set(s) of near replicates (typically the intended replicates)**
- 3. For each set of near replicates:**
 - 1. Calculate the independent variable center of mass for the set**
 - 2. Use the regression model to correct each point's observed response to the center of mass**
 - 3. Use these corrected values to estimate pure-error for each set**
- 4. Pool pure-error estimates, if there are multiple sets of near replicates**
- 5. Corrected values are only used for pure-error estimation**



Basic Visualization



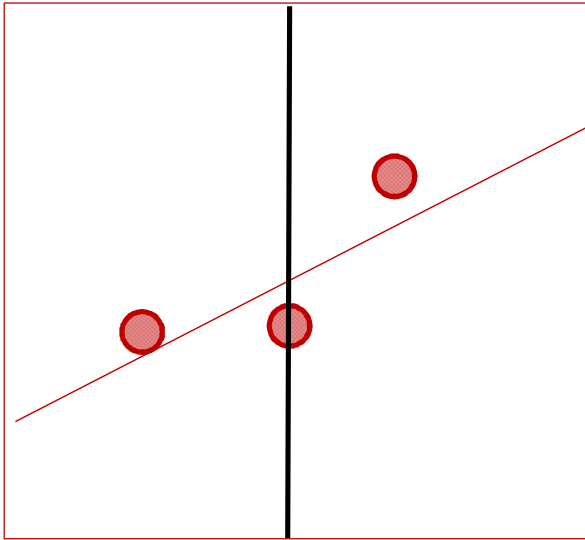
Center of Mass on X-axis

- Point on left will be adjusted upward by impact of slope difference from the center of mass
- Point on right will be adjusted downward by impact of slope difference from the center of mass
- Mid-point will be barely adjusted upward

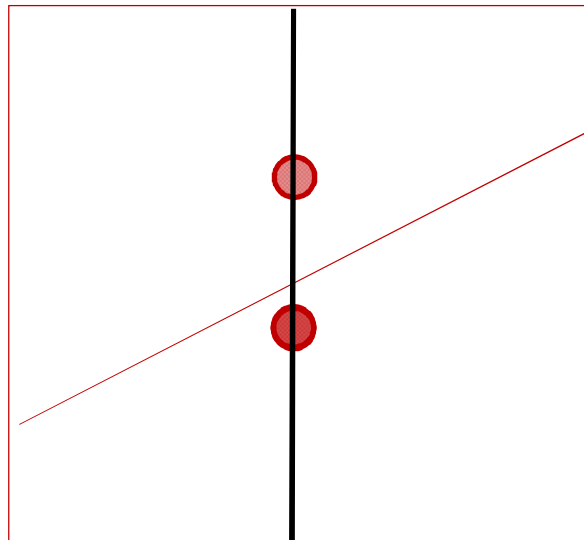
Visualizing the Strategies



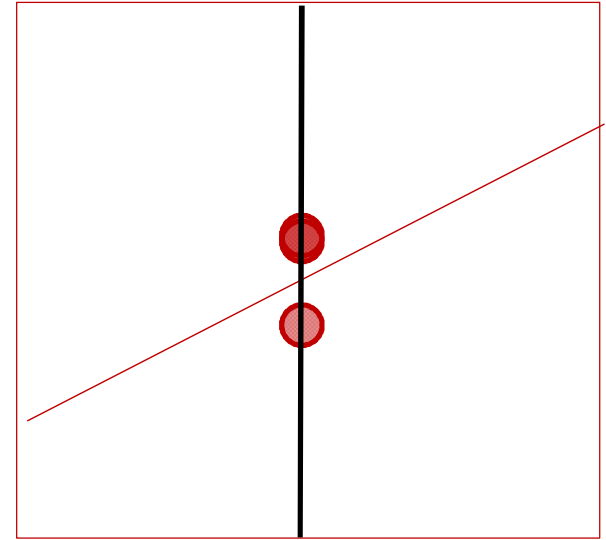
Near Replicate



Naive Approach
Pretend the Same X



Compensated
by Model to Same X





Simple Numerical Example

X	Y
1.0	15.0
1.8	19.0
2.0	20.0
2.2	21.0
3.0	25.0

True Model: $Y = 10 + 5X$ (without any simulated error)

Fitted Model: $Y = 10 + 5X$

The near replicates were three attempts to get $X=2.0$

Naïve estimate of PE variance = Variance (19, 20, 21) = 1

Center of mass of Near Replicates = Mean(1.8, 2.0, 2.2) = 2.0

X-offset = Center of Mass - X

Adjusted Amount = X offset * est. slope = X offset * 5

Adjusted Y = Y + Adjusted Amount

Corrected PE variance = Variance (20, 20, 20) = 0

X	Y	X Offset	Adjust Amount	Adjusted Y
1.8	19.0	0.2	1	20.0
2.0	20.0	0.0	0	20.0
2.2	21.0	-0.2	-1	20.0



Generalized Approach to Quasi Pure-Error

The basic idea is to identify a set of points that are within a defined neighborhood where they are effectively, but not precisely, replicates. Let $\mathbf{X}_{\text{nr}}$ be the $r \times p$ vector of the locations for this set of points, and let $\mathbf{y}_{\text{nr}}$ be the corresponding vector of observed responses. Note that $\mathbf{X}_{\text{nr}}$ has the form

$$\mathbf{X}_{\text{nr}} = \begin{bmatrix} \mathbf{x}'_{\text{nr},1} \\ \mathbf{x}'_{\text{nr},2} \\ \vdots \\ \mathbf{x}'_{\text{nr},r} \end{bmatrix}.$$

Let $\mathbf{x}'_0$ be the center of mass based on $\mathbf{X}_{\text{nr}}$, let $\mathbf{X}_0 = \mathbf{1}_r \mathbf{x}'_0$, and let $\mathbf{y}_{\text{corr}}$ be the vector of “corrected” responses, which uses the prediction model to adjust $\mathbf{y}_{\text{nr}}$ to the “expected” value if it were run at $\mathbf{x}_0$. As a result,

$$\mathbf{y}_{\text{corr}} = \mathbf{y}_{\text{nr}} - [\hat{\mathbf{y}}(\mathbf{X}_{\text{nr}}) - \hat{\mathbf{y}}(\mathbf{X}_0)].$$

where $\hat{\mathbf{y}}(\mathbf{X}_{\text{nr}}) = \mathbf{X}_{\text{nr}} \hat{\boldsymbol{\beta}}$ is the vector of predicted values at $\mathbf{X}_{\text{nr}}$ and $\hat{\mathbf{y}}(\mathbf{X}_0) = \mathbf{1}_r \hat{\mathbf{y}}(\mathbf{x}_0)$ is the predicted value at the center of mass. Let the prediction model be

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

where $\mathbf{y}$ is the $n \times 1$ vector of observed responses, $\mathbf{X}$ is the $n \times p$ model matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown coefficients, and $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of errors. We assume

$$\begin{aligned} \mathbf{E}(\boldsymbol{\epsilon}) &= \mathbf{0} \\ \text{var}(\boldsymbol{\epsilon}) &= \sigma^2 \mathbf{I}_n. \end{aligned}$$

Note that this correction assumes that the model is correct. The resulting standard ordinary least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Variance of Residuals at Near Replicates

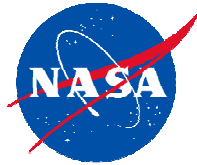


- Working through the derivation, it's clear that the approach is estimating the **variance of the near-replicate residuals**

$$\widehat{PE} = \text{var}(y_{nr} - \hat{y}_{nr})$$

- From a residual perspective, we assume that the variance of the near-replicate residuals is due to pure error only
 - True in the limit as the x-distance between the near-replicates goes to zero, near-replicates become exact replicates.
- Computing the variance of the near-replicate residuals is straightforward and practitioner friendly since they are available from the regression and design of experiments software packages.
- Using the residuals makes the method applicable to any model structure, regardless of model order, nonlinear, or non-parametric

Some Methodological Assumptions



- In the neighborhood of the near-replicates, we assume
 - True underlying function is locally smooth and relatively unaffected by lack-of-fit.
- While global lack-of-fit could be significant, in the neighborhood of intended exact replicates, small lack-of-fit would introduce bias and not inflate variance
 - Therefore, it would not significantly affect the pure-error estimate for practical situations
 - The greater the distance between the near-replicates, there is more potential influence of lack-of-fit.
- Simulation study probes the impacts of distance between the near-replicates (nearness) and lack-of-fit in the model.



Center Replicated, No Model LoF

Linear Model, No Model LOF, Replication at Center

Model: $Y = 100 + 20 \cdot X + \varepsilon$ where $\varepsilon \sim N(0, 1)$, $MSE = 1$

Simulated Near Replicate Variation in X

Range %	Naïve Variance	Corrected Variance
0%	0.9994	0.9994
2%	1.0427	1.0018
10%	2.0055	1.0009
20%	5.0073	0.9980
40%	17.0053	0.9870
100%	100.9540	0.9308
200%	400.7180	0.8365

Range % (variation in near replicates)
= 4 times the std. dev. of X divided by the DOE's range, expressed as a percentage

Naïve Variance:

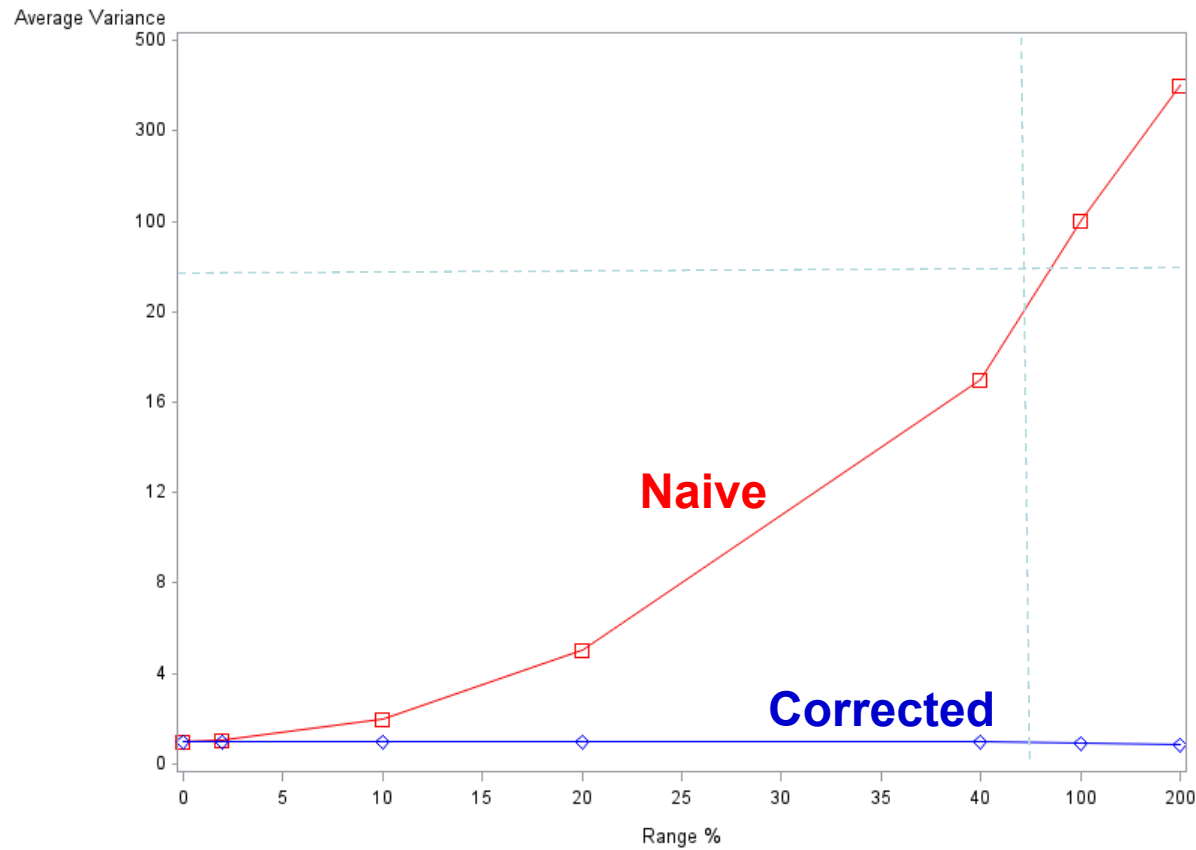
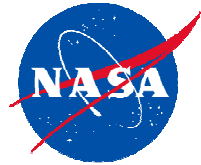
The variance computed by pretending the near replicates are true replicates.

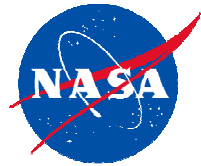
Corrected Variance

The variance after applying the correction methodology.

Variance close to 1 indicates best results

Center Replicated, No Model LoF





Practical Levels of Nearness

Std Dev X	Range %	Rational?
0	0%	NA
0.01	2%	Yes
0.05	10%	Yes
0.1	20%	Maybe
0.2	40%	Stretch
0.5	100%	No
1	200%	No

**One-dimensional cases simulations studied the full range
Subsequent simulations were restricted to reasonable rang
Considered method performance in its **realm of likely application.****

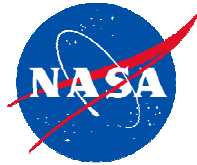
Center Replicated ($X = 0$), Model LOF



Nonlinear Model, Model LOF, Replication at Center		
Model: $Y = 100 + 20 * ((x+2)^{1.3}) + \epsilon$ where $\epsilon \sim N(0,1)$, $MSE=1$		
Fit Model: $Y = a + bX$		
Simulated Near Replicate Variation in X		
Range %	Naïve Variance	Corrected Variance
0%	1.0018	1.0018
2%	1.1057	1.0018
10%	3.5682	1.0012
20%	11.2497	1.0002
40%	41.9150	1.0102
100%	254.2128	1.4557
200%	919.8681	4.1044

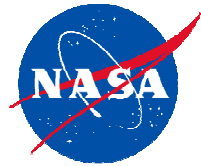
Worked well, but eventual bias in correction outside rational application range

Stressing the Correction Methodology



- **The prior testing was performed under moderately harsh test conditions (high model R-square, e.g., 99.5%). Increasing the simulated model MSE (lowered R-square) improves the relative method performance for all illustrated scenarios.**
- **Higher degrees-of-freedom also helps method performance in all tested scenarios.**
- **Our goal is to probe the methodology's behavior. Good, bad or approximate can be judged by individuals depending on application.**

Jumping to Three Dimensions: The Design Layouts



**2³ factorial with replicated
center (3x)**

X1	X2	X3
-1	-1	-1
-1	-1	1
-1	1	-1
-1	1	1
1	-1	-1
1	-1	1
1	1	-1
1	1	1
0	0	0
0	0	0
0	0	0

**2³ factorial with (1, 1, 1)
replicated corner (3x) and center**

X1	X2	X3
-1	-1	-1
-1	-1	1
-1	1	-1
-1	1	1
1	-1	-1
1	-1	1
1	1	-1
0	0	0
1	1	1
1	1	1
1	1	1



Center Replicated, No Model LOF

Case: True Model is Multivariate, No Model LOF, Replication at Center				
Simulated Model: $Y = 50 + 20x_1 + 5x_2 - 5x_3 - 5x_1x_2 + \varepsilon$ where $\varepsilon \sim N(0,1)$, $MSE=1$				
Fitted Model: $Y = a + bx_1 + cx_2 + dx_3 + ex_1x_2$				
Simulated Near Replicate Variation in X1				
Range %	Naive Variance	Corrected Variance	Naive Power	Corrected Power
0%	1.0018	1.0018	0.050	0.050
2%	1.0427	1.0018	0.048	0.050
5%	1.2539	1.0017	0.044	0.050
10%	2.0055	1.0015	0.031	0.050
15%	3.2567	1.0011	0.020	0.050
20%	5.0073	1.0005	0.014	0.050
40%	17.0053	0.9968	0.004	0.050

- With no LOF the methodology remains robust in the presence of more dimensionality
- Appears that the method was unaffected by increased dimensionality.



Center Replicated, Model LOF

Case: True Model is Multivariate, Model LOF, Replication at Centerpoint

Simulated Model: $Y = 50 + 20x_1 + 5x_2 - 5x_3 - 5x_1x_2 + 5x_1x_1 + \varepsilon$ where $\varepsilon \sim N(0,1)$, $MSE=1$

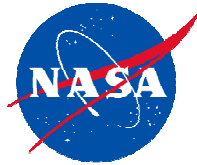
Fitted Model: $Y = a + bx_1 + cx_2 + dx_3 + ex_1x_2$ (Incomplete Model)

Simulated Near Replicate Variation in X_1

Range %	Naive Variance	Corrected Variance	Naive Power	Corrected Power
0%	0.9998	0.9998	0.521	0.521
2%	1.0378	0.9974	0.508	0.521
5%	1.2453	0.9940	0.448	0.525
10%	1.9959	0.9966	0.316	0.527
15%	3.2509	1.0059	0.210	0.526
20%	4.9967	1.0025	0.146	0.533
40%	17.0095	1.1106	0.053	0.538

- **With LOF the center point replicates still worked well**
- **Begins to deteriorate at 40%, however it pushing the practical boundaries**

Simulation Summary



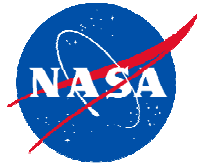
Naïve: PE estimates inflated by how near replicate variation is transmitted

1. Transmission through uncorrected model (often substantial)
2. Transmission of locally differential LOF (often smaller)

Corrected: PE estimates inflated by how near replicate variation is transmitted

2. Transmission of locally differential LOF (often smaller)

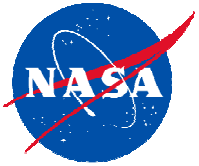
Concluding Remarks on Near Replicate RSM Technique



Demonstrated Proposed Correction Methodology is:

- **Easy to implement, even with standard RSM/DOE software**
- **Is robust with no model LoF is present**
- **Fairly robust with moderate LoF is present**
- **Fairly robust to moderate variation in near replicate values**

Even when imperfect, we recommend that this correction methodology be added to your RSM toolkit.



Questions?