Using Design of Experiments Methods for Efficient Modeling & Simulation

***Tom Donnelly**, PhD, CAP*
Principal Systems Engineer
JMP Defense & Aerospace Team
Tom.Donnelly@jmp.com
302-489-9291

DATAWorks 2024
DOE Short Course
April 16
IDA

jmp STATISTICAL DISCOVERY FROM SAS

Abstract

By sequentially running High Performance Computer (HPC) simulations in Design of Experiments (DOE) blocks, one can quickly increase the accuracy of fast surrogate models. When sufficiently precise surrogates have been obtained, HPC resources can sooner be made available to other Modeling & Simulation (M&S) projects.

Outline

- Links to Examples & Resources
- Why Use Design of Experiments (DOE) for M&S?
- Contrast Traditional DOE vs. Space-Filling DOE
  - Use JMP to create both types of designs
- Efficient M&S Using DOE – 3 Examples – 1 in detail
  - Sequential Traditional DOE
  - **Space-Filling DOE Case Study**
    - Create Design
    - Explore Data Visually
    - Regression and Machine Learning Models
    - Model Comparison – Honest Assessment – Train, Validate(Tune), Test
  - Sequential Space-filling DOE

Takeaways

DOE
Part

- JMP provides both *"Traditional"* (2 or 3-level) RSM, Factorial, OA, DSD & GOSSD, AND *Space-Filling* (many level) Latin HyperCube (LHC), Fast Flexible Filling (FFF), and more designs for computer experimentation
- After running and analyzing a DOE for a computer simulation, one can provide an *instantaneous answer for any new scenario*

Surrogate
Modeling
Part

- Able to deliver an *interactive trade-space analysis*
- Honest assessment approach of splitting data into TVT subsets makes applying machine learning methods *robust to overfitting*
- *Robust Machine Learning Strategy*
  - 1) *Bootstrap Forest* – fast even with many Xs – unlikely to miss factors
  - 2) *Neural Network* – often most flexible and best predictor
  - 3) *Penalized Regression* – often more interpretable + confidence intervals

Download & Recording

- 16 Factors
- 50,000 unique cases
- Each 1,000 times
- 50 Million Simulations
- Neural Network for Surrogate Models

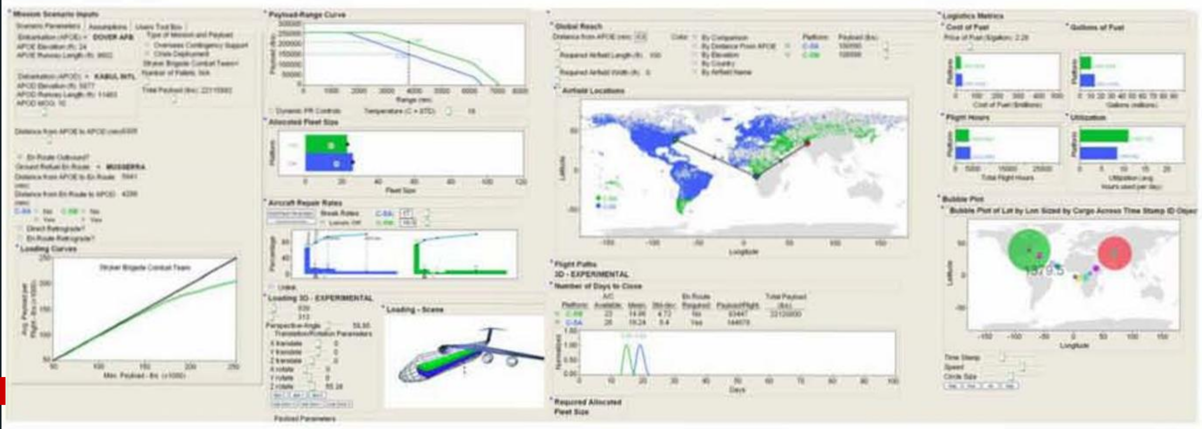
1.6 Comparative Assessment and Decision Support System for Strategic Military Airlift Capability

Comparative Assessment and Decision Support System for Strategic Military Airlift Capability

John Salmon, Curtis Iwata, Dimitri Mavris and Neil Weston
Georgia Institute of Technology
john.salmon@asdl.gatech.edu, curtis.iwata@asdl.gatech.edu, dimitri.mavris@aerospace.gatech.edu, neil.weston@ae.gatech.edu
Philip Fahringer
Lockheed Martin Company
philip.fahringer@lmco.com

ABSTRACT

The Lockheed Martin Aeronautics Company has been awarded several programs to modernize the aging C-5 military transport fleet. In order to ensure its continuation amidst budget cuts, it was important to engage the decision makers by providing an environment to analyze the benefits of the modernization program. This paper describes an interface that allows the user to change inputs such as the scenario airfields, take-off conditions, and reliability characteristics. The underlying logistics surrogate model was generated using data from a discrete-event simulation. Various visualizations, such as intercontinental flight paths illustrated in 3D, have been created to aid the user in analyzing scenarios and performing comparative assessments for various output logistics metrics. The capability to rapidly and dynamically evaluate and compare scenarios was developed enabling real-time strategy exploration and trade-offs.



View 5-min Video

Figure 2. Strategic Airlift Comparison Tool Layout

Download Document

https://ntrs.nasa.gov/search.jsp?R=20110012110

https://community.jmp.com/t5/US-Federal-Government-JMP-Users/VIDEO-of-C-5-Decision-Support-Tool-Dashboard/ta-p/34364

Download

Georgia Tech Library

SMARTech Home / Georgia Tech Theses and Dissertations / Georgia Tech Theses and Dissertations / View Item

A framework for the optimization of doctrine and systems in Army Air Defense units using predictive models of stochastic computer simulations

View/Open

WADE-DISSERTATION-2017.pdf (65.95Mb)

This thesis presents a new methodology that can be used to address large-scale raids made up of different types of Theater Ballistic Missiles (TBMs) and Cruise Missiles (CMs) that attempt to overwhelm the Air Defense Artillery (ADA) systems at a particular location. The primary focus will be on how existing ADA systems can adjust their tactics in order to minimize the damage caused by threats that are not shot down and impact friendly forces. Nearly all the literature to date optimizes systems and tactics to reduce the number of leakers – threats not shot down – that impact the ground. However, simply counting the number of leakers does not adequately describe the effects to friendly forces. Instead, the first part of this thesis combines existing methods for external ballistics, concrete penetration, explosive cratering, and weapon blast and fragmentation damage in order to create an integrated program that can describe the damage to an airfield runway, infrastructure, and parked aircraft. The second part of this thesis focuses on modeling the ADA missile engagements using an accredited Department of Defense ADA simulation model called the Extended Air Defense Simulation (EADSIM). Both the airfield damage model and ADA simulation have run times ranging from minutes to hours.

2017 Dissertation by

LTC Brian Wade

(Former Director, TRAC-MTRY, Naval Postgraduate School)

https://smartech.gatech.edu/handle/1853/58275

Additional Links…

- 2020 Military Operations Research Society (MORS) Modeling & Simulation (M&S) Community of Practice (CoP) webcast of full Tutorial…

  https://community.jmp.com/t5/US-Federal-Government-JMP-Users/Efficient-M-amp-S-Using-Sequential-DOE/ta-p/352332

- JMP Discovery Summit 2020 webcast of *JMP BEAST Mode: Boundary Exploration through Adaptive Sampling Techniques*

  https://community.jmp.com/t5/Discovery-Summit-Americas-2020/JMP-BEAST-Mode-Boundary-Exploration-through-Adaptive-Sampling/ta-p/281514?_ga=2.7898738.1561325551.1600970078-320379481.1600180249

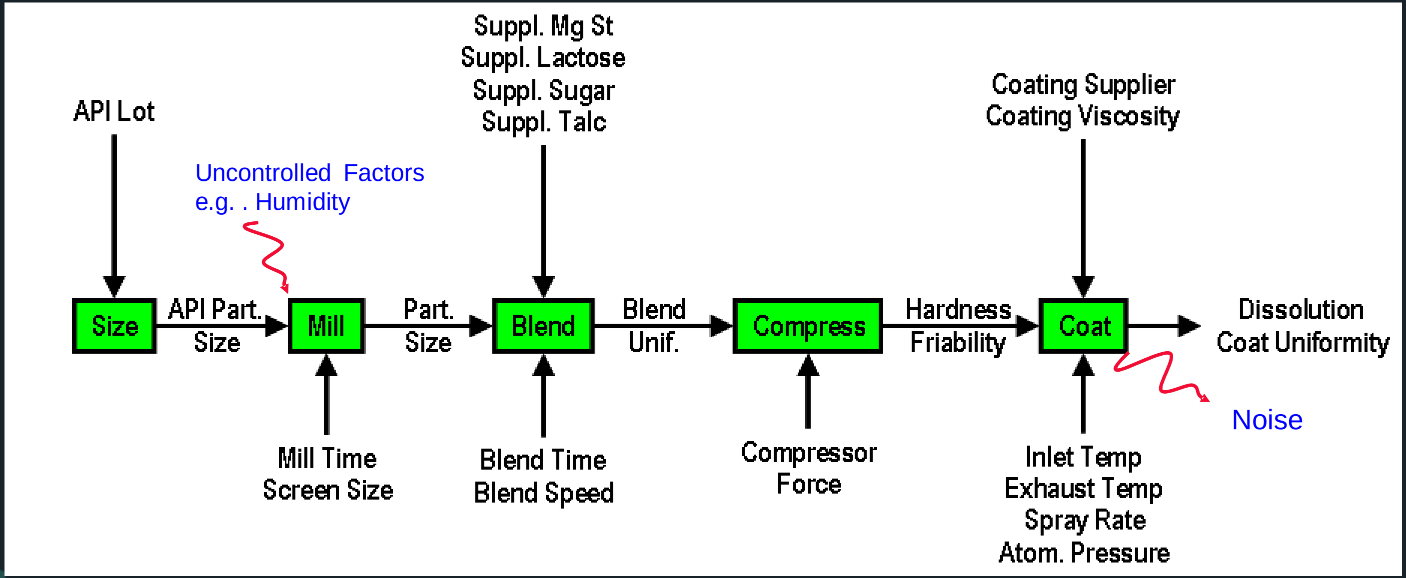
Why Use Design of Experiments Methods with Simulation Experiments?

Quicker answers, lower costs, solve bigger problems

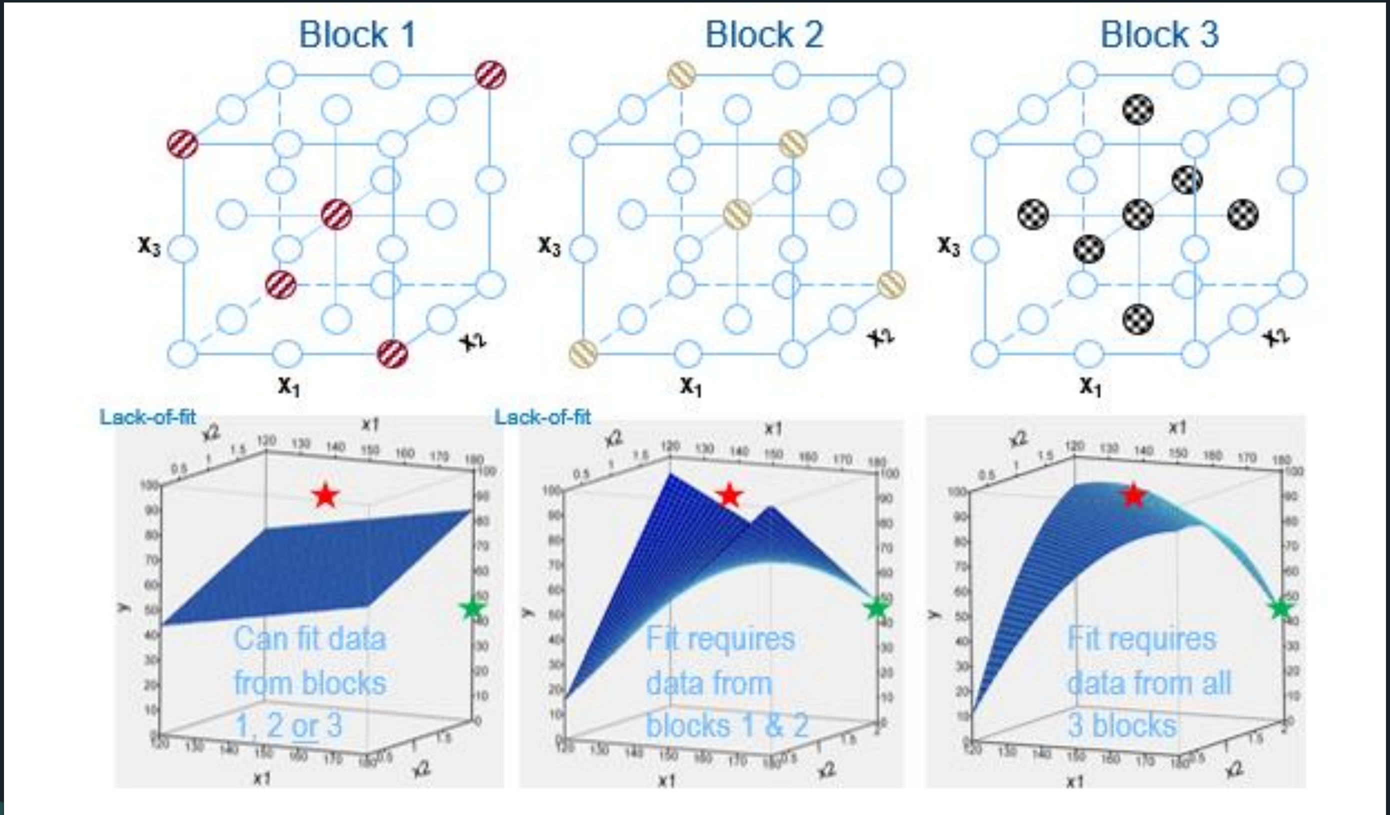
- Obtain a ***fast surrogate model*** of the ***slow running simulation***
  - Individual simulations can run for hours, days, weeks
    - Computational Fluid Dynamics (CFD) or Simulation runs in real-time
  - Numbers of factors can be very large (100+)
  - Numbers of simulations needed can be large (thousands in many cases)
  - Simulations can be stochastic requiring many replications
- Surrogate model is a ***fast approximation*** of the simulation
  - more rapidly answer “what if?” questions – ***Instantaneous answer for any “NEW” scenario!***
  - do sensitivity analysis of the control factors
  - optimize multiple responses and make trade-offs
- By running sequences of designs one can be as ***cost effective as possible*** & ***run no more trials than are needed*** to get a useful answer
- By running efficient subsets of all possible combinations (DOEs), one can – for the same resources and constraints – ***solve bigger problems***

Classic Definition of DOE

- Purposeful control of the inputs (factors) in such a way as to deduce their relationships (if any) with the output (responses).



Response Surface DOE in a Nutshell



Design of Experiments

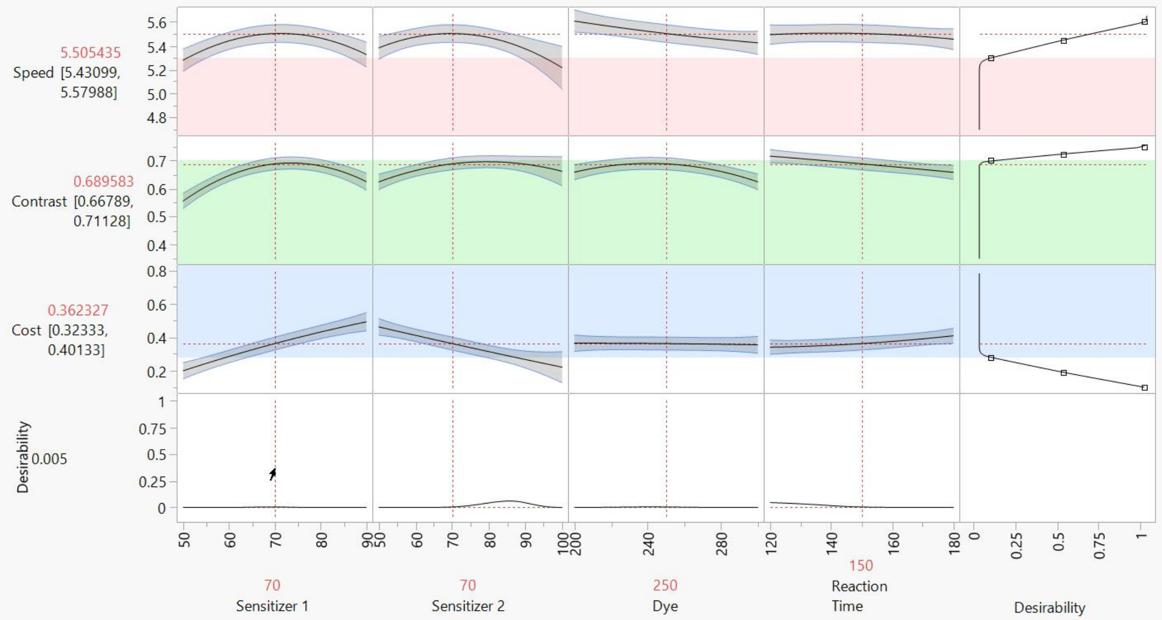
Derive maximum information from fewest tests – yielding "interactive" trade-off and optimization

Photo_Cost27 - JMP Pro [3]

File Edit Tables Rows Cols DOE Analyze Graph Tools Add-Ins View Window Help

7/4 Cols
27/0

	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
1	50	50	250	120	5.36	0.616	0.198
2	50	50	200	180	5.39	0.537	0.175
3	90	70	200	120	5.31	0.623	0.447
4	50	90	200	150	5.13	0.431	0.177
5	70	70	250	180	5.37	0.643	0.445
6	50	90	300	120	4.79	0.375	0.231
7	90	90	200	180	5.45	0.626	0.471
8	90	50	250	150	5.00	0.470	0.670
9	50	50	300	150	5.22	0.478	0.283
10	70	90	200	120	5.41	0.668	0.226
11	90	90	250	120	5.33	0.734	0.310
12	50	50	250	120	5.32	0.574	0.257
13	70	50	200	150	5.49	0.596	0.456
14	50	70	250	180	5.22	0.558	0.166
15	70	70	250	150	5.57	0.689	0.390
16	90	90	300	150	5.26	0.653	0.226
17	70	70	250	150	5.47	0.688	0.356
18	70	70	300	120	5.42	0.657	0.337
19	50	70	200	120	5.43	0.518	0.222
20	50	50	300	150	5.15	0.505	0.287
21	90	70	200	120	5.33	0.661	0.457
22	50	90	300	120	4.97	0.411	0.191
23	90	50	300	120	5.09	0.492	0.588
24	90	50	300	180	5.03	0.358	0.733
25	70	70	250	150	5.59	0.707	0.318
26	70	90	300	180	5.25	0.605	0.290
27	50	90	200	150	5.24	0.476	0.177



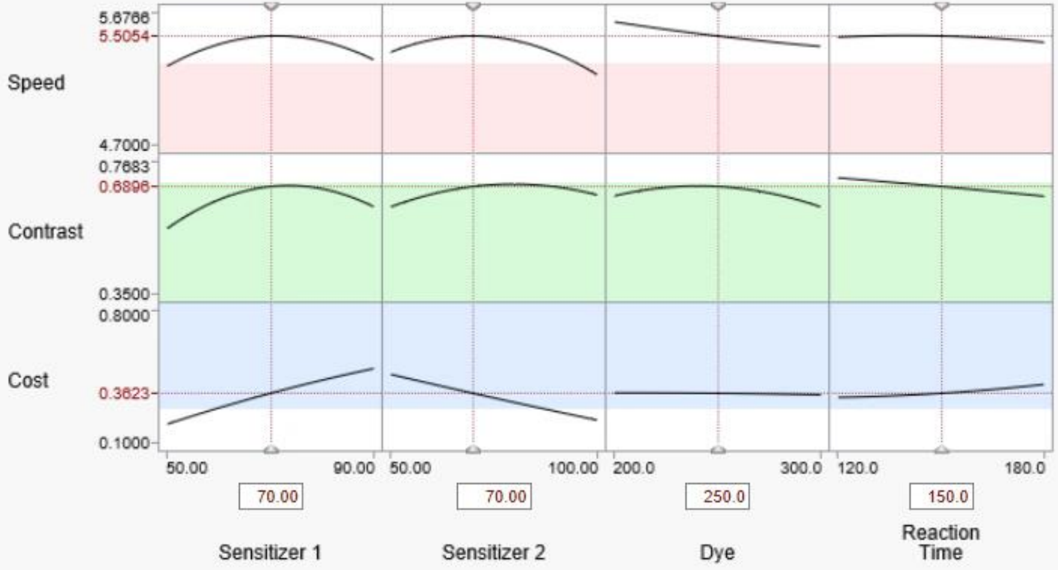
https://devlive16.jmp.com/packages/byzlbyzbc_GGvpFg2wS0k-4

JMP STATISTICAL DISCOVERY FROM SAS

Share results

ON JMP PUBLIC OR JMP LIVE

Prediction Profiler



Remembered Settings

	Setting	Sensitizer 1	Sensitizer 2	Dye	Reaction Time	Speed	Contrast	Cost
Apply	Equal Importance Opt	80.753574	91.269729	250.57625	120	5.3542877	0.7466933	0.2504014
Apply	Mid Point Settings	70	70	250	150	5.5054353	0.6895831	0.3623274
Apply	Cost 6X Speed & Contrast	84.016038	93.725925	283.02514	120	5.2902084	0.72549	0.1991539
Apply	Opt Spd3X-Cntr1X-Cost6X	81.958309	90.706277	286.82246	120	5.3269582	0.7177857	0.2211116

View optimizations on your phone. Scan the QR code to launch browser, then use finger to interact with the Prediction Profiler and to “Apply” saved settings.

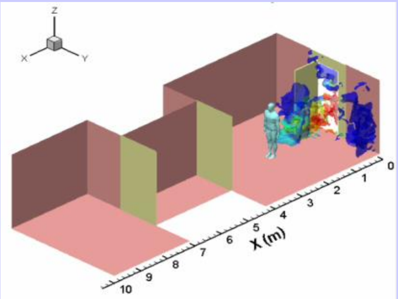
Three Main Types of Simulations

- Complex science-based models such as computational fluid dynamics (CFD) codes. Solves differential equations at boundary conditions of millions of cells. Most computer intensive. Usually, deterministic (same answer each run).
- Agent-Based Models (ABM) use rules for individual agents to study the behaviors and interaction of agents in a system. Usually, stochastic (different answer due to randomness). Repeat many times to see “average” behavior.
- Discrete Event Simulation (DES) models a system as a sequence of events. Usually, stochastic. Repeat many times to see “average” behavior.

Long Running Physics-Based Simulations

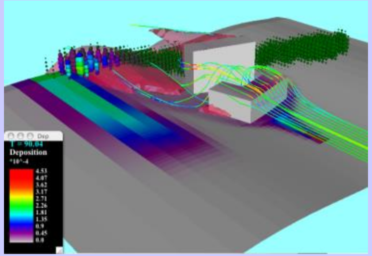
Detailed Physics Models can require a great deal of runtime to generate a short period of simulation time.

Computational Fluid Dynamics (CFD) Models



Developed for Interior
Moving Man in Simulation
8M cells
10 Seconds of Simulation
64 CPUs – 4K slower
12 Hours of Runtime

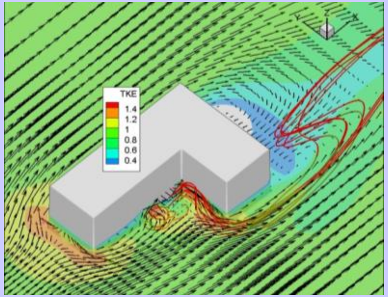
Detailed Ingress/Egress, Internal Airflow and Convection



Developed for Exterior
Stationary Grids
1.5M Cells
30 Seconds of Simulation
Single CPU – 20K slower
7 Days of Runtime

External CW Deposition/ Evaporation, Vegetation, Solar Heating

Lagrangian-Particle

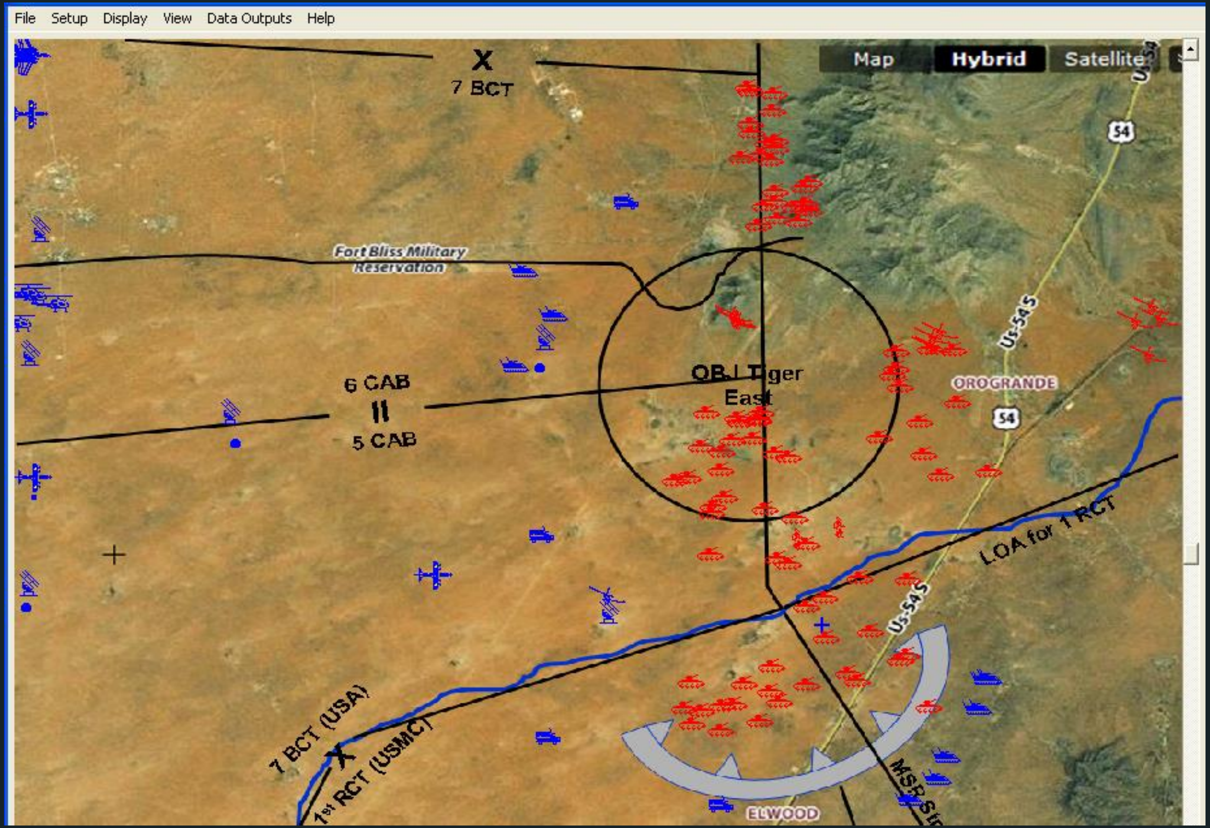


Developed for Exterior
Stationary Grids
TBD Cells
Min-Hours of Simulation
Single CPU
Minutes-Days of Runtime

Speed, Flexibility, More User Friendly, V&V

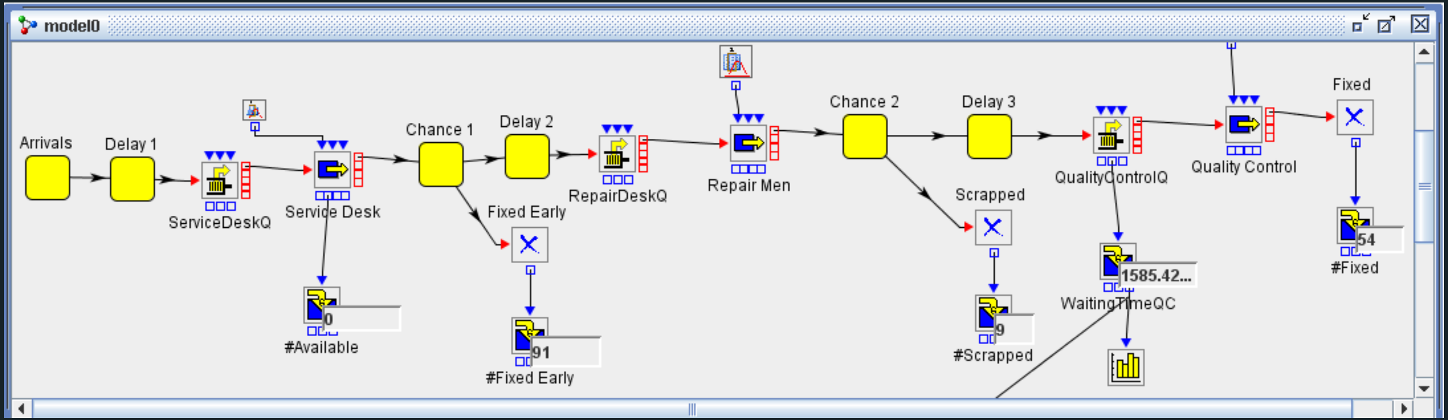
Stochastic SImulations with Many Replicates

Agent Based Simulations



Stochastic Simulations with Many Replicates

Discrete Event Simulations

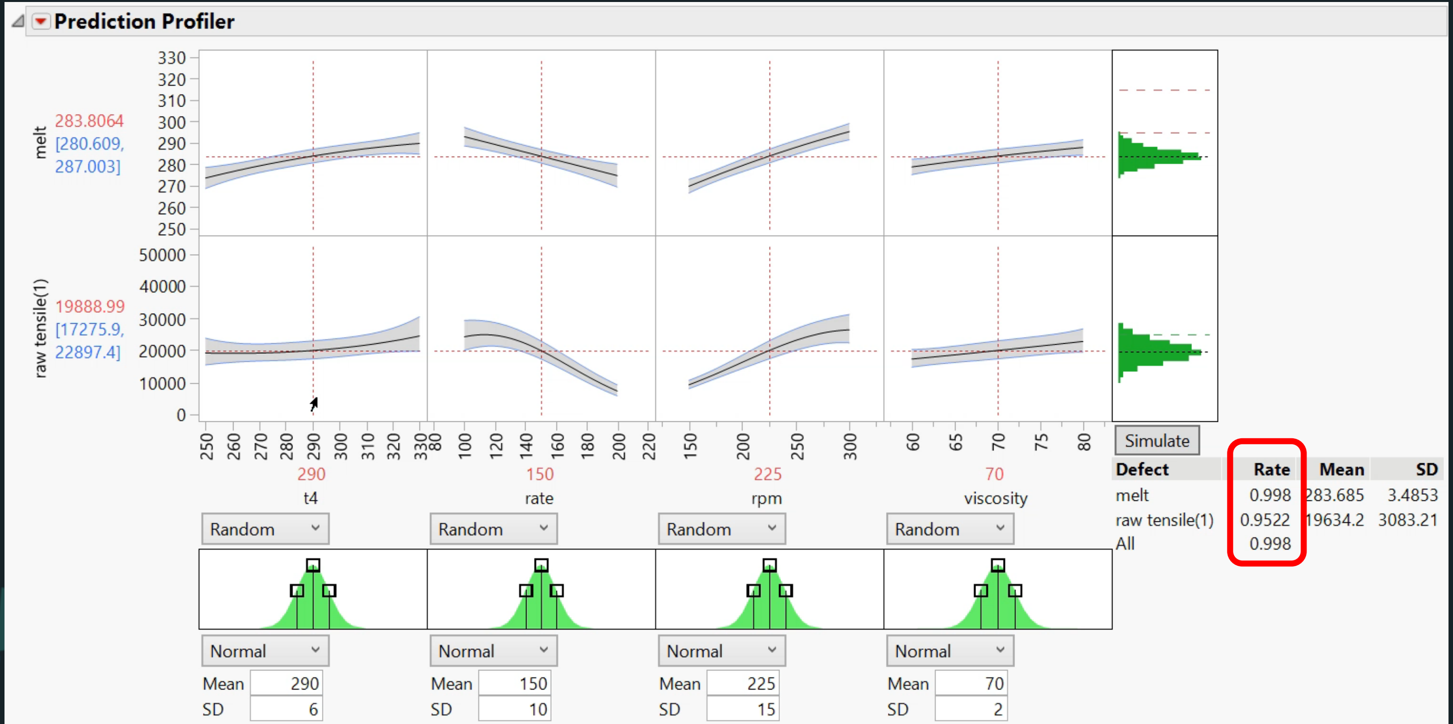


experiment0

Point...	StartT...	EndTi...	Num...	Num...	Num...	Repli...	Num...	AvgUt...	AvgW...	AvgUti...	AvgUt...	AvgW...	AvgW...
point 1	0	2,700	1	1	3	5	55.8	98.1...	640....	32.90...	22.7...	1.681...	0.11...
point 2	0	2,700	3	2	1	5	105.6	61.8...	3.65...	16.39...	67.8...	0.0	16.7...
point 3	0	2,700	2	3	1	5	100.2	88.3...	84.8...	10.92...	67.8...	0.0	16.7...
point 4	0	2,700	2	1	3	5	100.6	88.5...	97.3...	32.90...	22.7...	1.681...	0.11...
point 5	0	2,700	2	1	1	5	100.2	88.3...	84.6...	32.78...	67.8...	0.233...	16.7...
point 6	0	2,700	3	1	2	5	105.8	61.9...	8.69...	32.90...	34.1...	1.382...	0.83...
point 7	0	2,700	2	2	2	5	100.4	88.4...	97.9...	16.47...	34.1...	0.020...	0.83...
point 8	0	2,700	2	2	3	5	100.6	88.5...	98.4...	16.46...	22.7...	0.094...	0.11...
point 9	0	2,700	1	1	1	5	55.8	98.1...	621....	32.78...	67.8...	0.233...	16.7...
point ...	0	2,700	3	3	3	5	105.8	61.9...	9.32...	10.97...	22.7...	0.001...	0.11...
point ...	0	2,700	1	3	2	5	55.8	98.1...	641....	10.98...	34.1...	4.305...	0.83...
point ...	0	2,700	1	2	1	5	55.8	98.1...	621....	16.39...	67.8...	0.0	16.7...

Assess Uncertainty in Surrogate Model Predictions for a Deterministic Simulation with No Replications

For non-stochastic simulations for which a surrogate model has been created, Monte Carlo simulations can be run using assumed distributions for inputs to better assess transmitted variation about the model point estimate.



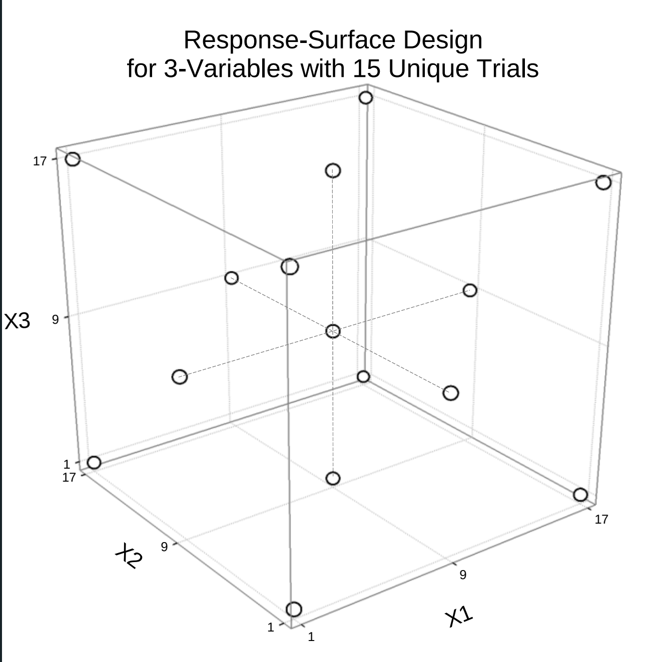
Defect	Rate	Mean	SD
melt	0.998	283.685	3.4853
raw tensile(1)	0.9522	19634.2	3083.21
All	0.998		

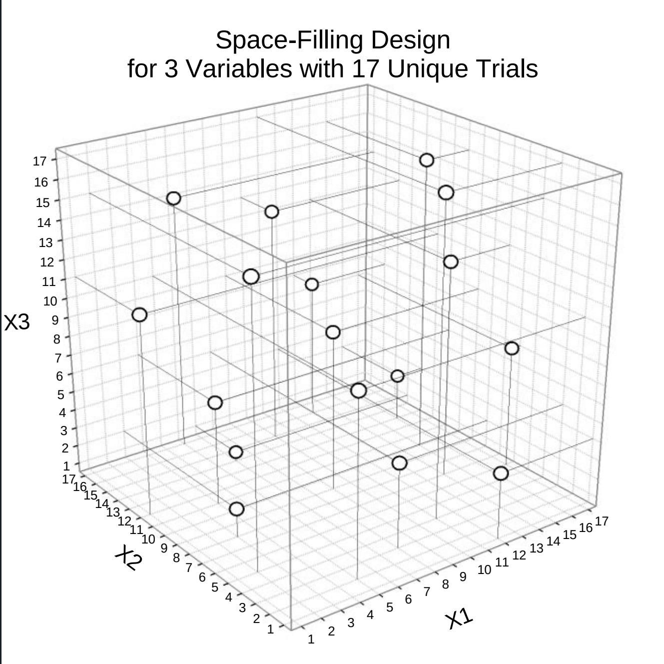
Two Classes of Designs for Two Types of Surrogate Modeling of Simulations

- *Traditional RSM/Factorial/OA/DSD/GOSSD* designs for modeling with "low-order polynomials" with continuous (quantitative) and categorical (qualitative) factors and usually stochastic responses – more applicable for DES and ABM simulations
  - Designs can be sequentially constructed to support increasingly complex models
  - Example featured here reanalyzes a simulation case matrix in which all combinations of 6 factor settings were originally run – a total of 648 = 6 X 3 X 3 X 3 X 2 X 2 times.

- *Space-filling* designs primarily for use with continuous (but also categorical) factors AND usually non-stochastic/deterministic responses – more applicable with CFD
  - These designs can support "Gaussian Process" or "Kriging" "spatial regression" analysis – an interpolation technique, as well as linear regression – an approximation method
  - Data for all above designs can be modeled with JMP's full range of regression, neural network, and partition (regression & classification tree) methods.

How are Space-Filling Designs Different from Traditional Designs?





Rather than emphasizing high leverage trials ("corners") for a simple polynomial model, space-filling designs "spread" their trials more uniformly through the space to better capture the local complexities of the simulation model.

Use JMP to Make Both Types of Designs

- Orthogonal Arrays (OA) and Nearly Orthogonal Arrays (NOA) for any combination of levels* (Or use FF, PB, DSD, RSM, GOSSD)
  - JMP 14 and earlier: DOE > Screening Designs
  - JMP 15 and later: DOE > Classical > Two-level Screening > Screening Design

- Space-Filling Designs : DOE > Special Purpose > Space Filling Design
  - Latin HyperCube (LHC) – most popular choice of top 6 designs for continuous factors
  - Fast Flexible Filling (FFF) – handles real-world problems that LHC cannot
    - constraints
    - categorical factors
    - augmentation

*Lekivetz, R. (2006). ***A New Algorithm for Obtaining Mixed-Level Orthogonal and Nearly-Orthogonal Arrays***, M.Sc. Thesis, Dept. of Statistics and Actuarial Science, Simon Fraser University.

Traditional Designs for Polynomial Modeling

- I used to say, “If a textbook fractional-factorial, orthogonal array or response-surface design is available, then use it.”

  Now I say, “If Definitive Screening Design is available, then use it.”
- Textbooks and web site catalogs do not always contain designs for categorical variables with:
  - all combinations of mixed numbers of levels (e.g., 3, 4, 5, and 21)
  - large numbers of levels for variables (e.g., 5+)
- Algebraic (Orthogonal Array) and algorithmic (D-optimal) computer generated designs can often be used
  - Orthogonal Arrays (and Nearly Orthogonal Arrays) are good at yielding analysis with un-confounded estimates of the “main effects” when variables have many different levels
  - D-optimal designs are good for adding on the fewest additional trials to support higher order “interaction” terms in the model

Sequential Designs

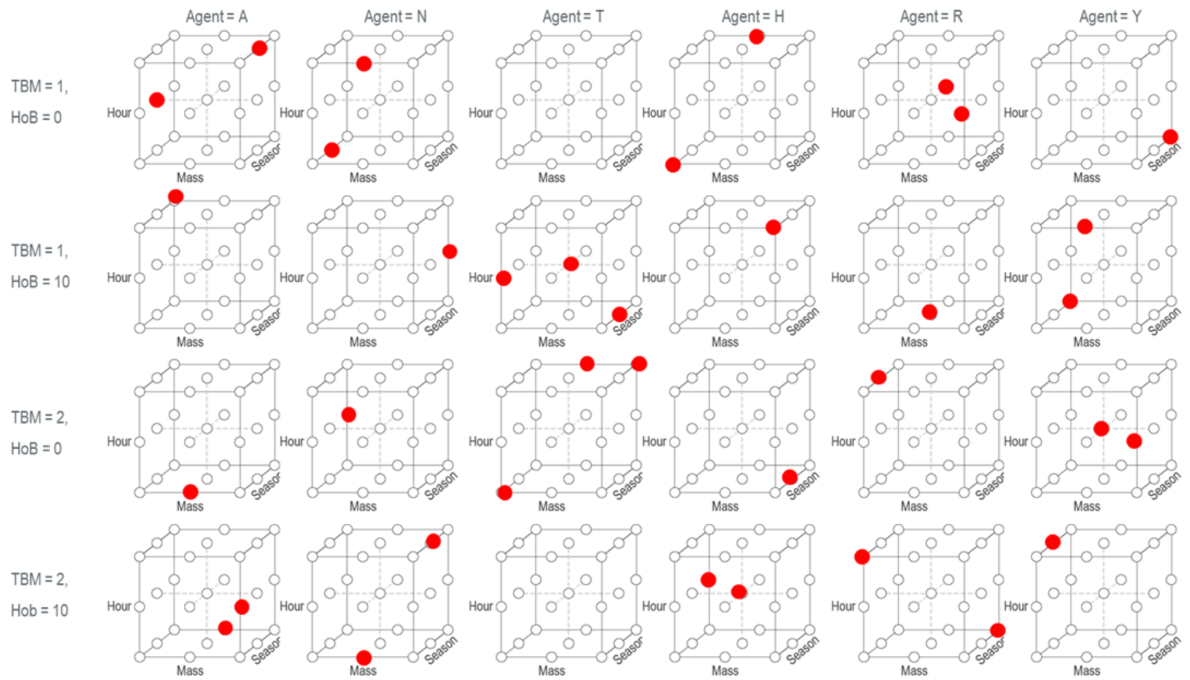
- Simulation experiments – Sequential designs are easily employed because "restricted randomization" is not an issue
  - Many simulations are deterministic
  - Even if stochastic (random), correlation with unknown factors is not possible
  - All factors are generally just as easy to change
  - Can still inexpensively add a blocking variable to test if "the code has been changed!"
- Real experiments – The issue of "restricted randomization" does arise making sequential experimentation a bit more complicated – but still possible to employ
  - Groups of trials run at different (even widely spaced) periods of time
    - Addressed using a *blocking* factor
  - Sometimes there are factors that are harder to change than others, e.g. *Oven Temperature*
    - Addressed using *split-plot* designs

Case Matrix as Used in Study of the Observed Response "Probability of Casualty" (PCAS)

Variable	# Levels	Levels
Agent Codes	6	A, N, T, H, R, Y (categorical)
Season	3	Winter, Summer, Spring/Fall (categorical)
Time of Attack (Hour)	3	0500, 1200, 2200 Local Time (continuous)
No. of TBMs & Spread Radius	2	1 TBM & 1 m, 2 TBMs & 1000 m (categorical)
Mass (relative)	3	1.00, 1.57, 2.00 (continuous)
Height of Burst (HoB)	2	0, 10 m (continuous)
Total Cases	648	

OA and NOA Designs

- Six categorical factors
  - X1 has 6-levels
  - X2, X3, & X4 have 3-levels
  - X5 & X6 have 2-levels



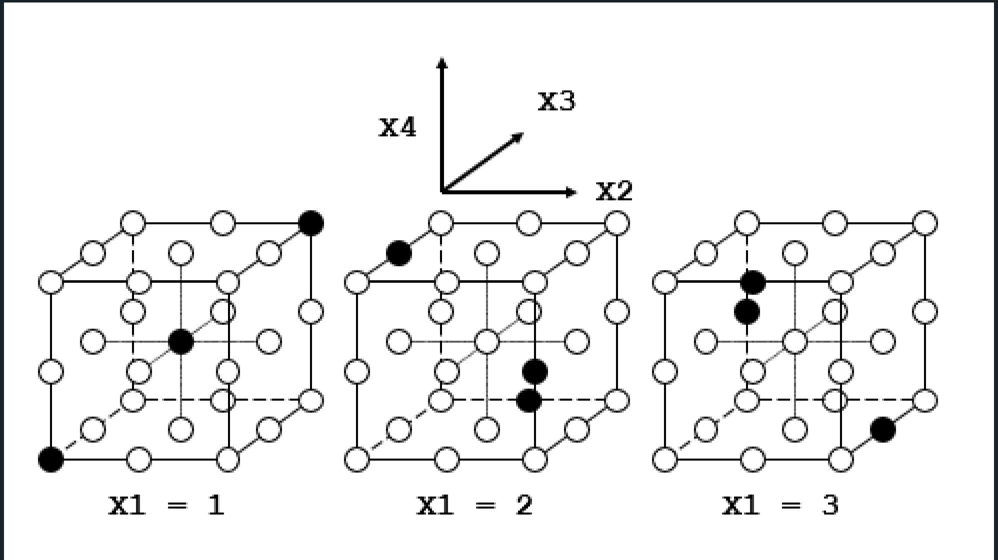
Red Dots Mark the 36 Trials (an Orthogonal Array) Analyzed for Stage 1 50

- 18, 36, 108, 324, 648 trials
  - Compare Distributions
  - Compare Scatterplot Matrices
  - Compare Color Maps of Correlations

36 (5.6%) of All 648 Possible Combinations of Settings for 6 Variables (6 X 3 X 3 X 3 X 2 X 2)

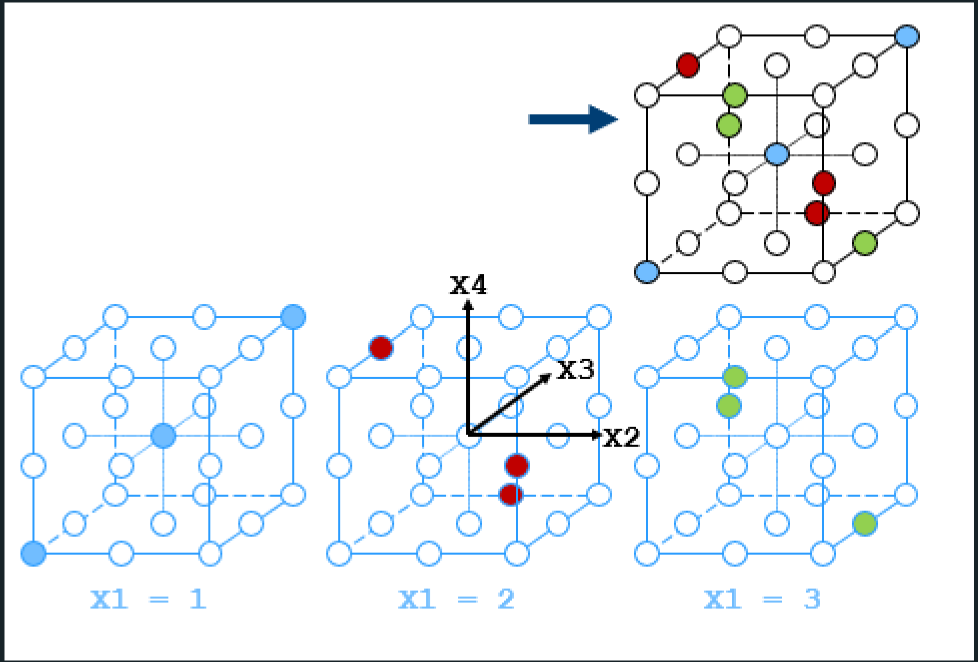
Locations of Trials for a 4-variable, 9-trial Orthogonal Array Design

x1	x2	x3	x4
1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	1	2
3	1	3	2
3	2	1	3
3	3	2	1

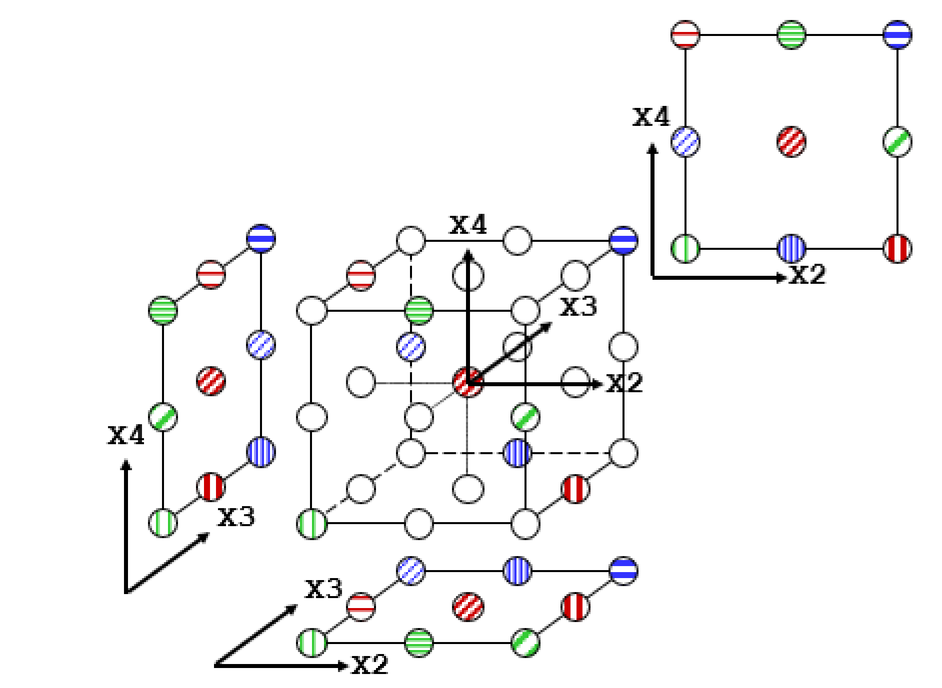


Delete **X1** and View Locations of Trials for a 3-Variable OA9 Design

X1	X2	X3	X4
1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	1	2
3	1	3	2
3	2	1	3
3	3	2	1



Projection of Trial Locations for a 3-variable OA9 Design for All Pairs of Variables



All projections have 9 unique trials that can be used to fit a 2-variable quadratic model with 6 terms

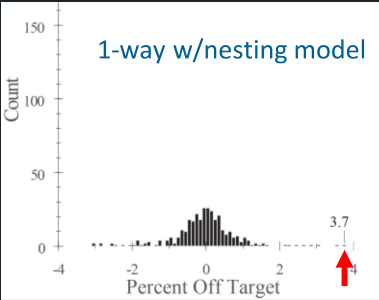
Four Stage Design Sequence

Stage 1	Stage 2	Stage 3	Stage 4
36 Total Simulations	**108** Total Simulations	**324** Total Simulations	ALL **648** Simulations
Design 1, 36 trials	Design 1, 36 trials	Design 1, 36 trials	Design 1, 36 trials
	Design 2, 72 trials	Design 2, 72 trials	Design 2, 72 trials
		Design 3, 216 trials	Design 3, 216 trials
Main effects only for ALL variables + some 2-way interactions	Stage 1 effects plus all 2-way interactions + some 3-way interactions	Stage 2 effects plus all 3-way interactions	Stage 3 effects plus ALL remaining 4-way, 5-way and 6-way interactions
5.6% of 648	16.7% of 648	50% of 648	Design 4, 324 trials NOTE: Length of this green box should be longer than shown

324 trials in Design 4 used as checkpoints for Designs 1, 2 & 3 →

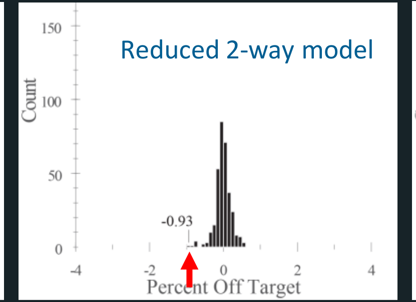
“Factor Sparsity” and “Effect Heredity” Used to Enhance Model Complexity

36 trials (5.6%)



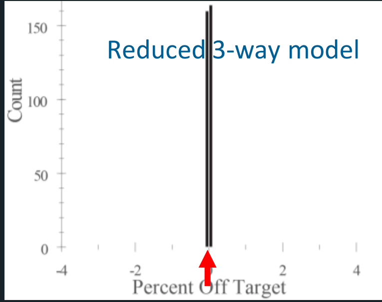
Worst Case = 3.7%
Half of Cases < 0.37%

108 trials (16.7%)

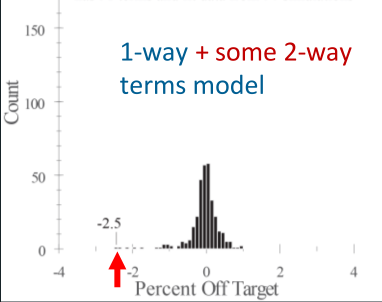


Worst Case = -0.93%
Half of Cases < 0.11%

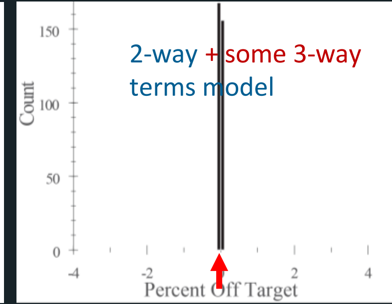
324 trials (50%)



Worst Case = -0.0081%
Half of Cases < 0.0007%



Worst Case = -2.5%
Half of Cases < 0.16%



Worst Case = -0.0251%
Half of Cases < 0.0010%

Histograms showing % off target for predicting 324 checkpoint trials.

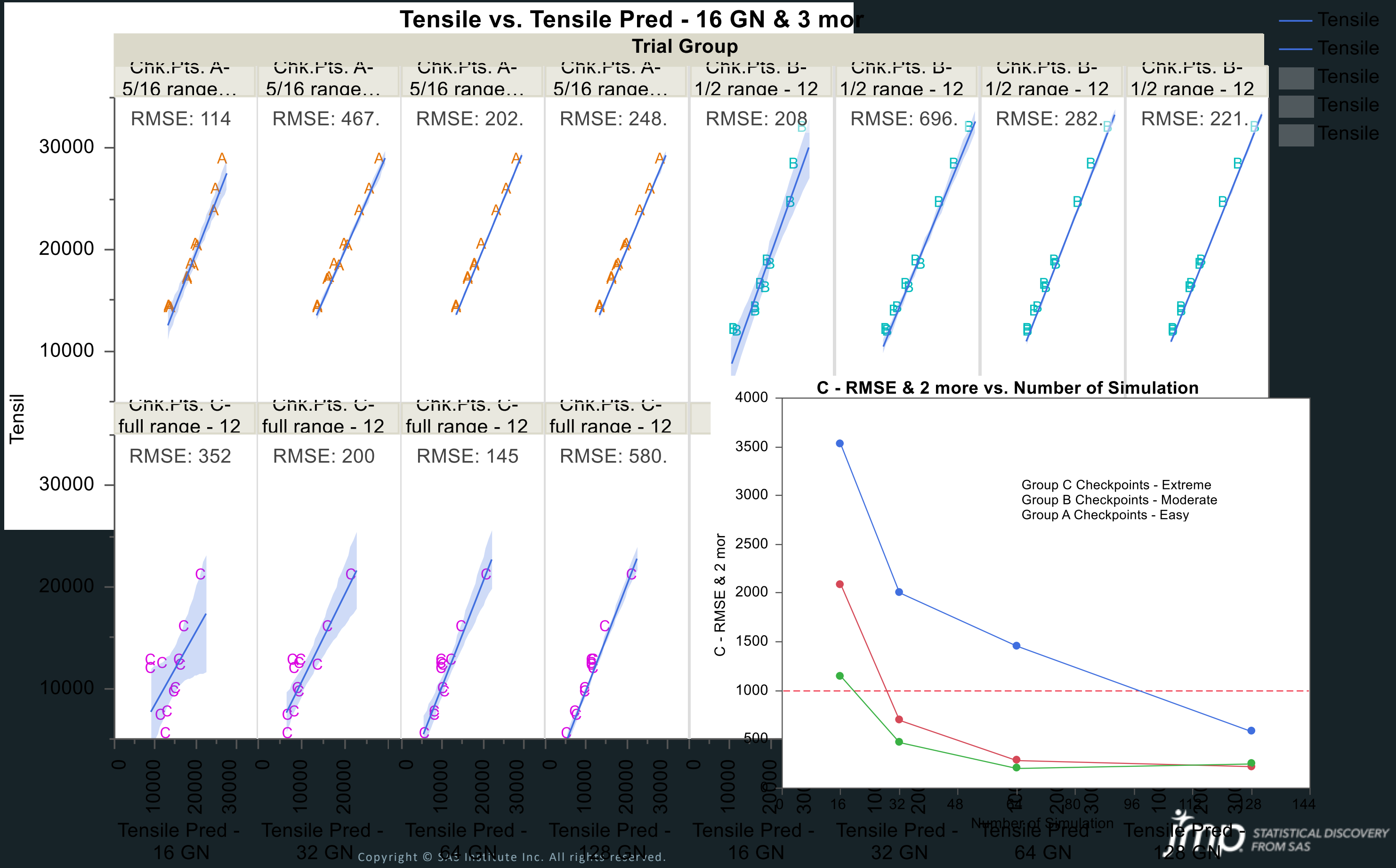
Factor Sparsity states only a few variables will be active in a factorial DOE

Effect Heredity states significant interactions will only occur if at least one parent is active
(See Wu & Hamada, p. 112)

Oct. 1, 2007 visit by Profs. Wu & Joseph of GA Tech ISyE

Plots of Actual vs. Predicted (Simulation vs. Surrogate) by Checkpoint Group for 4 Stages of Analysis of NLHD

Checkpoint Groups A & B show diminishing return in prediction improvement for running past stage 3



jmp STATISTICAL DISCOVERY FROM SAS

Go to JMP

- OA and Custom Designs with 18, 36, and 108 trials
  - Compare balance, color maps, and power

29 Simulations Run – 17 trial LHC design used to create surrogate model & 12 PB used as checkpoints

17-trial Orthogonal Latin Hypercube (OLH) space-filling design settings used for creating the surrogate model

17 levels for each factor

12-trial Plackett-Burman screening design settings used as checkpoints – half just inside and half just outside design boundary (convex hull)

Trial	Time of Day	Temperature	Wind Speed	Wind Direction	Relative Humidity	Cloud Cover	
1	505	37	5.3	247.5	30	0.92	
2	165	13	5.6	281.25	10	0.32	
3	250	19	1.7	225	60	0.8	
4	335	25	2.9	360	55	0.14	
5	1100	35	3.5	202.5	35	0.02	- Min
6	1440	15	3.2	326.25	15	0.74	
7	930	11	6.2	236.25	80	0.44	
8	845	33	5	348.75	75	0.62	
9	760	21	3.8	270	50	0.5	- Mid
10	1015	5	2.3	292.5	70	0.08	
11	1355	29	2	258.75	90	0.68	
12	1270	23	5.9	315	40	0.2	
13	1185	17	4.7	180	45	0.86	
14	420	7	4.1	337.5	65	0.98	- Max
15	80	27	4.4	213.75	85	0.26	
16	590	31	1.4	303.75	20	0.56	
17	675	9	2.6	191.25	25	0.38	
18	972.5	26	3.05	298.125	62.5	0.65	Inside
19	547.5	16	4.55	241.875	62.5	0.65	Outside
20	972.5	26	3.05	241.875	37.5	0.65	Outside
21	547.5	26	4.55	298.125	37.5	0.35	Outside
22	972.5	16	4.55	298.125	62.5	0.35	Inside
23	547.5	16	3.05	241.875	37.5	0.35	Inside
24	547.5	26	4.55	241.875	62.5	0.65	Outside
25	972.5	16	4.55	298.125	37.5	0.65	Inside
26	547.5	26	3.05	298.125	62.5	0.35	Inside
27	547.5	16	3.05	298.125	37.5	0.65	Outside
28	972.5	16	3.05	241.875	62.5	0.35	Outside
29	972.5	26	4.55	241.875	37.5	0.35	Inside

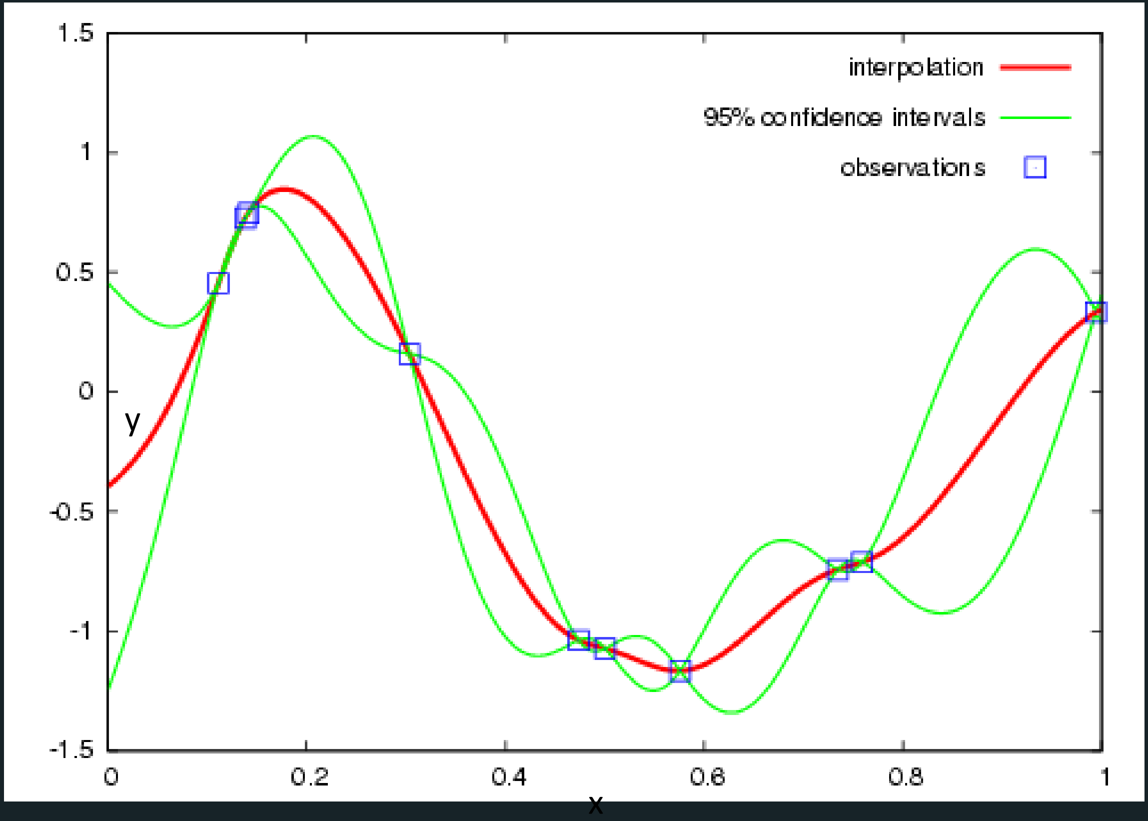
Go to JMP

- Latin HyperCube Design with 17 trials (or 33, 65, 129… trials) [Why $2^n + 1$ trials?]
  - Use 1 – 17 for range of X1
  - Use 2 – 98 for range of X2
  - Use 0 – 100 for range of X3

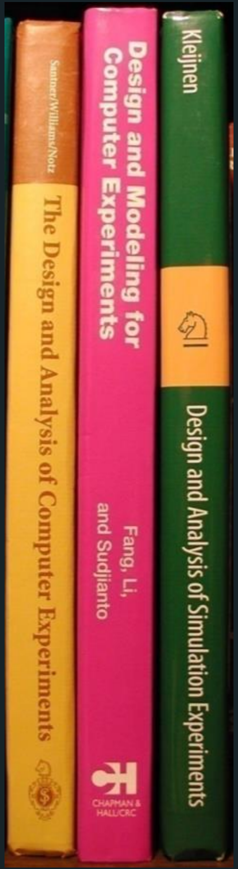
- DSD w/13t + FFF w/20t for 33 total vs. DSD w/17t + FFF w/16t for 33 total
  - Use 1 – 17 for range of X1
  - Use 2 – 98 for range of X2
  - Use 0 – 100 for range of X3

- GOSSD 192f in 108 trials (DSD is 2X @ 385 trials), 2^{192} = 6.277e57 corners
- Augment with FFF to 960 total trials
- If needed augment again

Gaussian Process/Kriging Fit in 1-D Showing Interpolation and Confidence Intervals



Seminal Paper on “Space-Filling” DOE for Computer Experiments



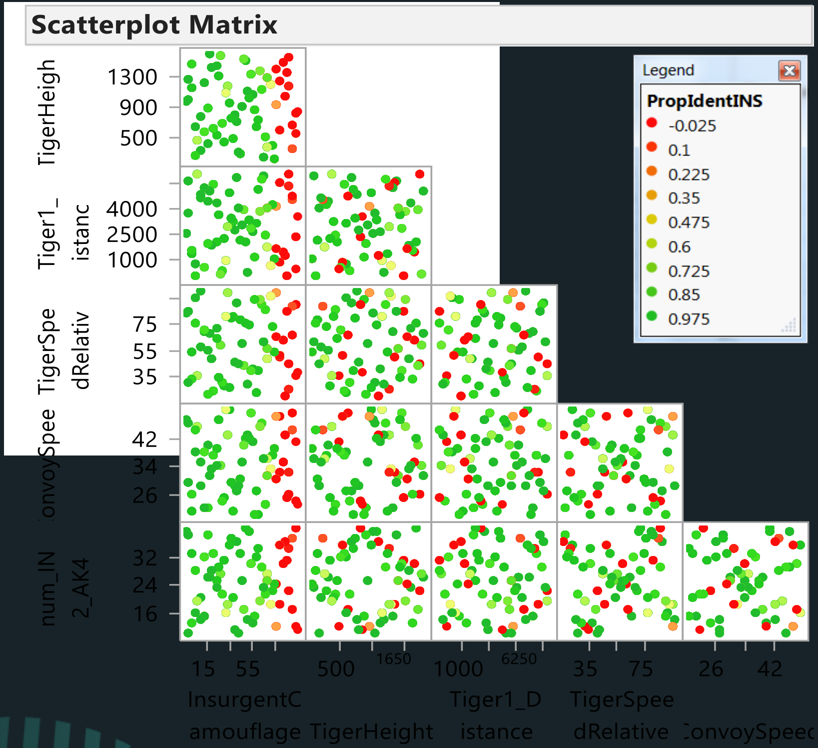
Design and Analysis of Computer Experiments
Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P.
Statistical Science 4. 409-423, 1989

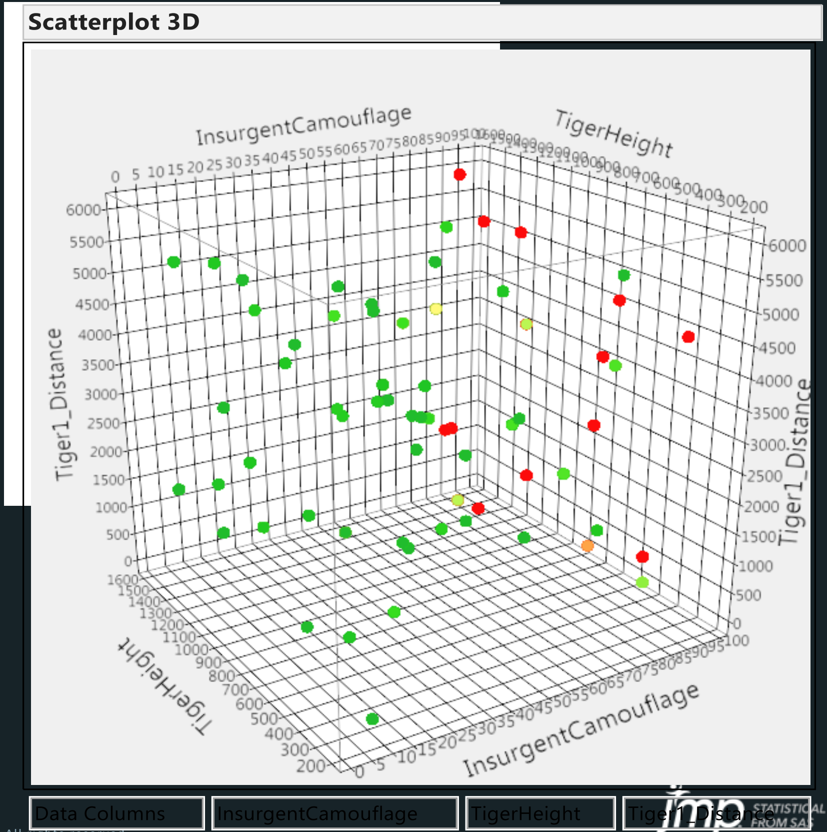
– Textbooks on this topic include:
  - Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, Springer, New York (2nd in 2018)
  - Fang, K. T., Li, R. Z., and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC Press, New York
  - Kleijnen, J. P. C. (2008), *DASE: design and analysis of simulation experiments*. Springer, New York. (2nd in 2015)

Surrogate Modeling of a Computer Simulation
Helicopter Surveillance – Identifying Insurgents

- 2009 International Data Farming Workshop - IDFW21, Lisbon, Portugal
- Largely German team (6 of 8) – their simulation
- 6500 simulations run overnight on cluster in Frankfurt
  - Space Filling Design of Experiments (DOE)
  - 65 unique combinations of 6 factors (each factor at 65 levels)
  - each case had 96 to 100 replications (lost a few)
- Response = Proportion of Insurgents Identified = *PropIdentINS* Data bounded between 0 and 1
- Create space-filling LHC designs, show augmentation with FFF
- Explore data visually first
- **Fit many different models – Regression and *Machine Learning*** using "Train, Validate (Tune), Test" subsets
- Compare Actual vs. Predicted for Test Subsets

Space-Filling DOE (Latin HyperCube) visualized with 2-D ScatterPlot MATRIX and 3-D ScatterPlot



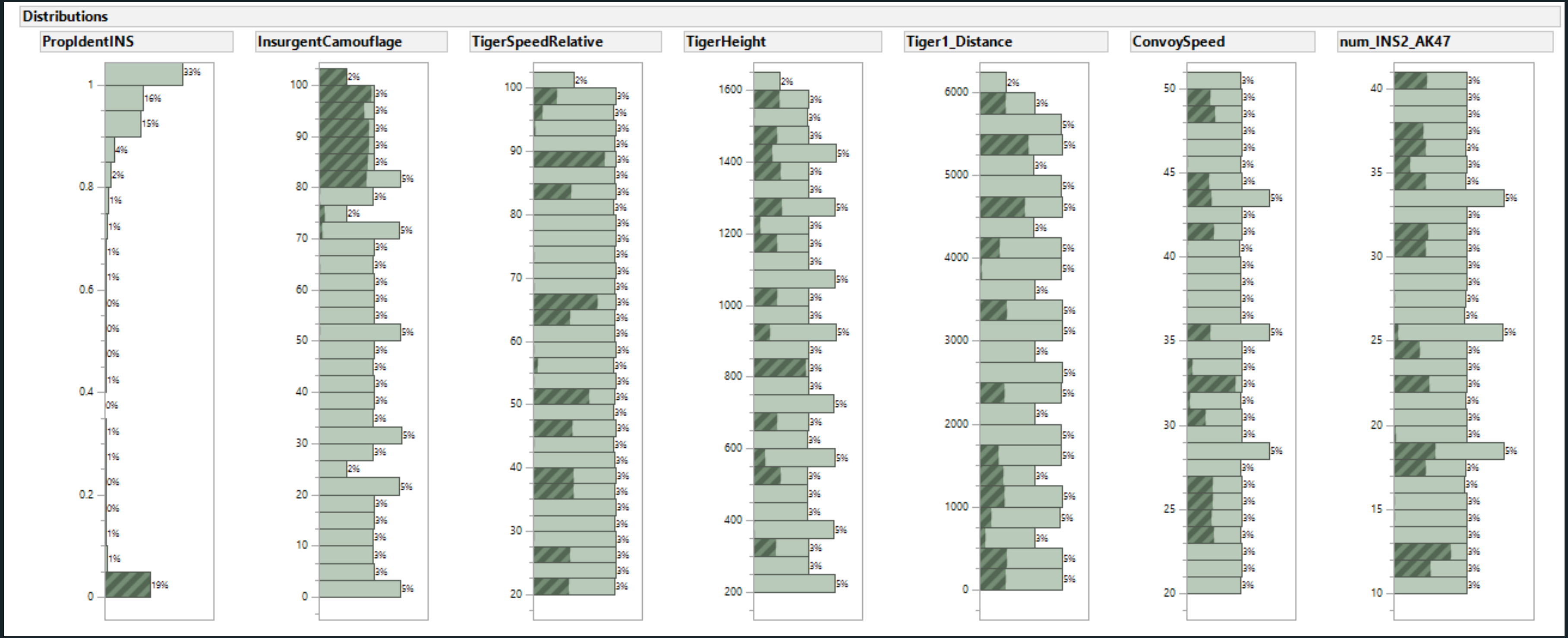


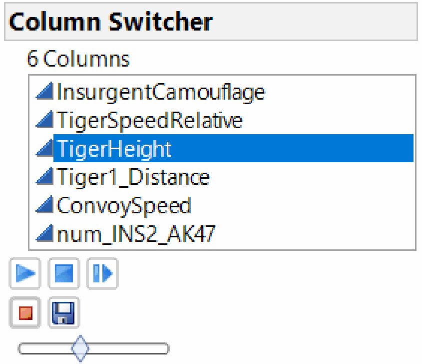
Go to JMP

- Create 65-trial LHC Design
  - Load factor ranges

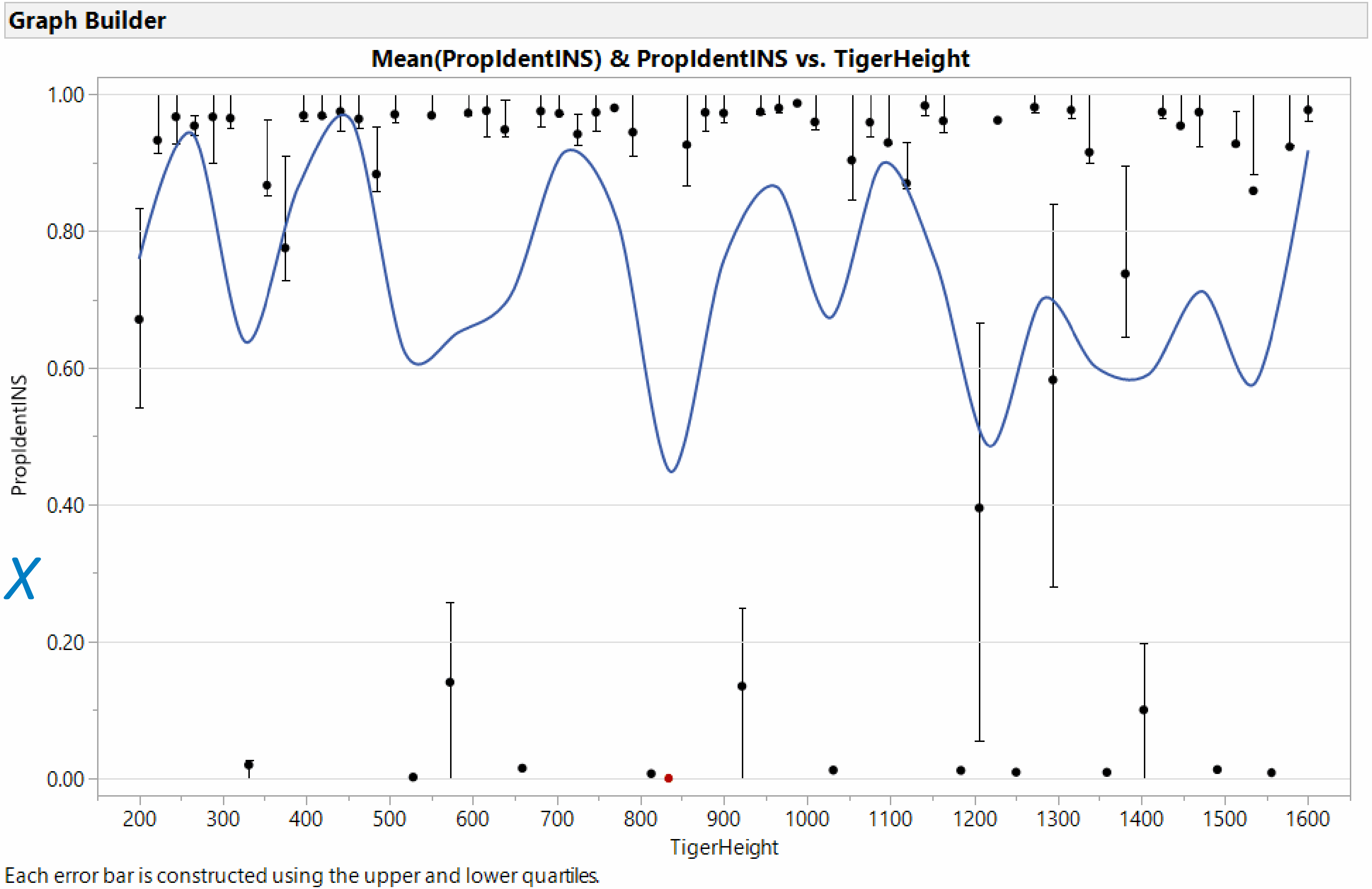
- Create FFF Design with
  - Constraint on X1 & X2
  - AND X6 is categorical

Distributions of 1 Response and 6 Factors

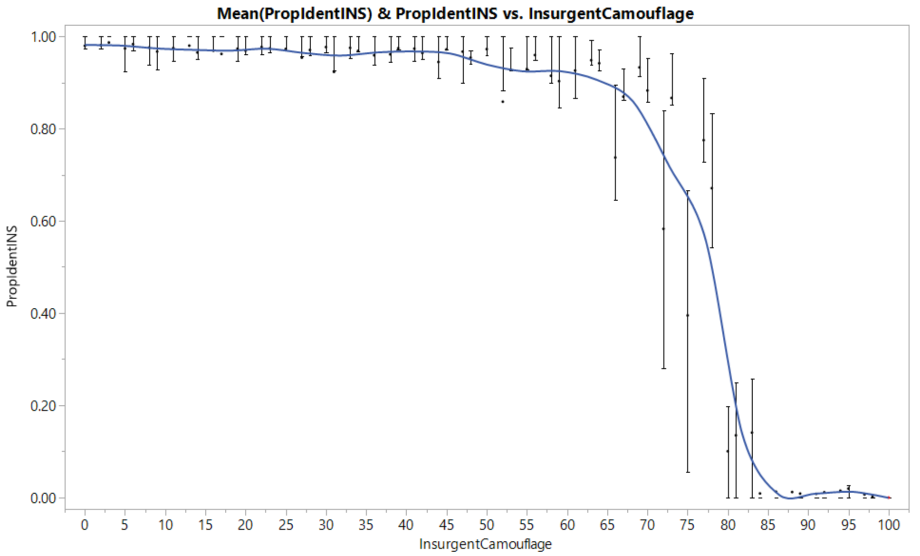


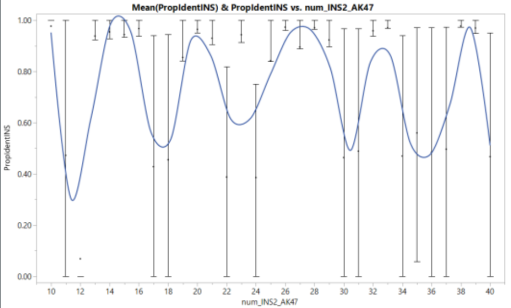


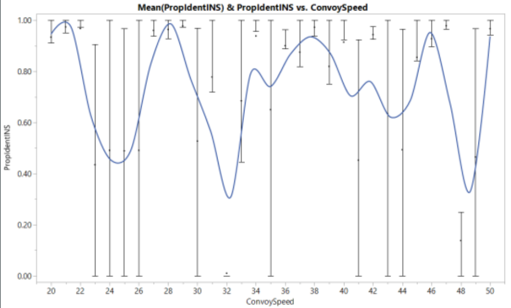
PropIdentINS vs. *X* for 6 Factors

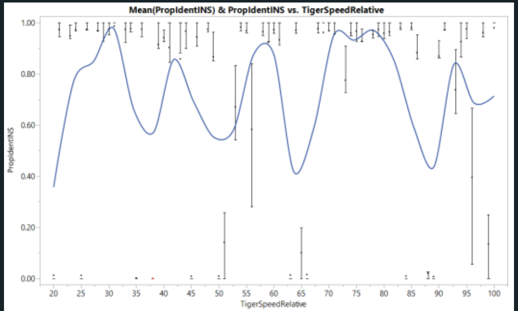


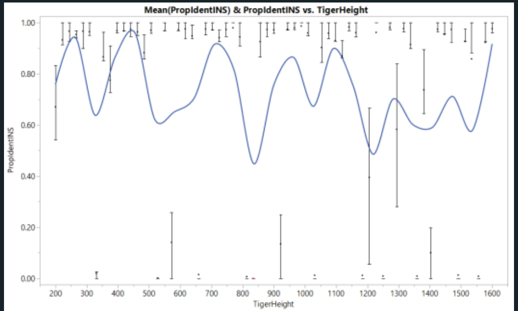
PropIdentINS vs. *X* for 6 Factors

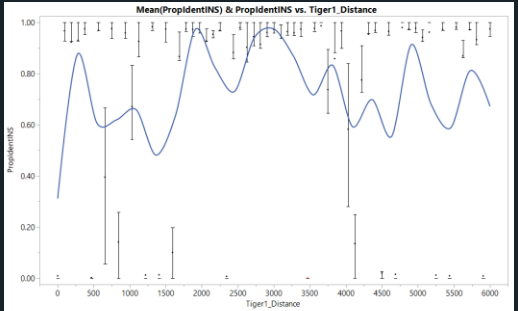




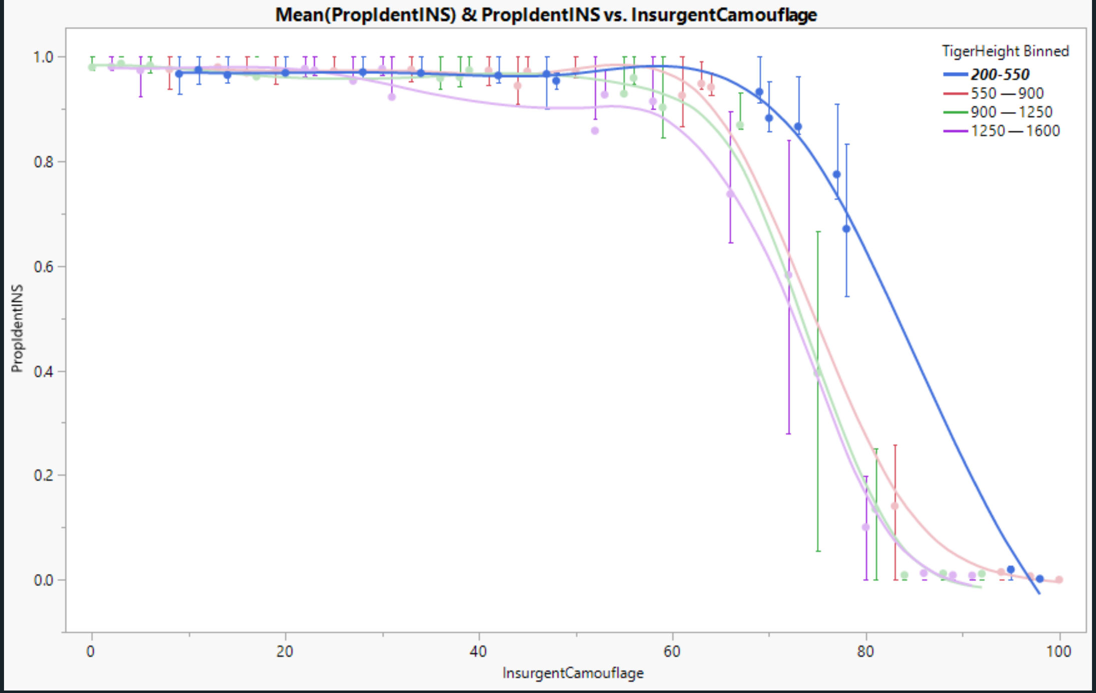


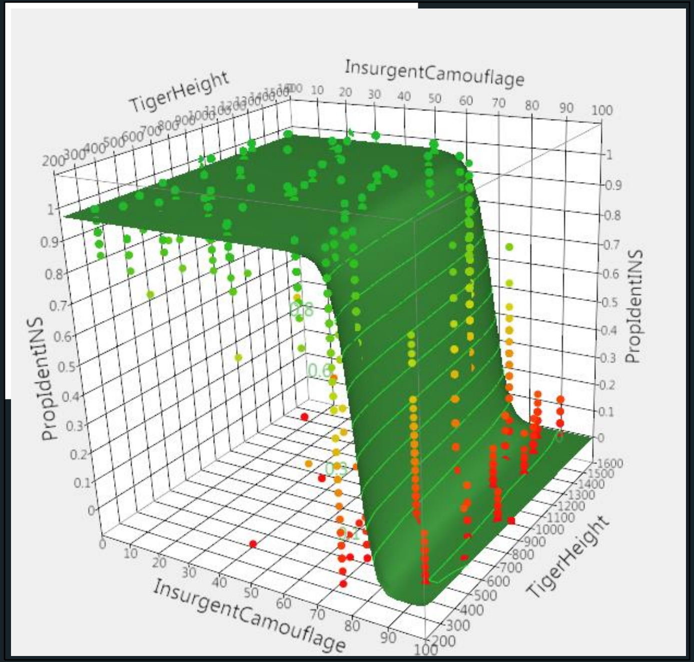






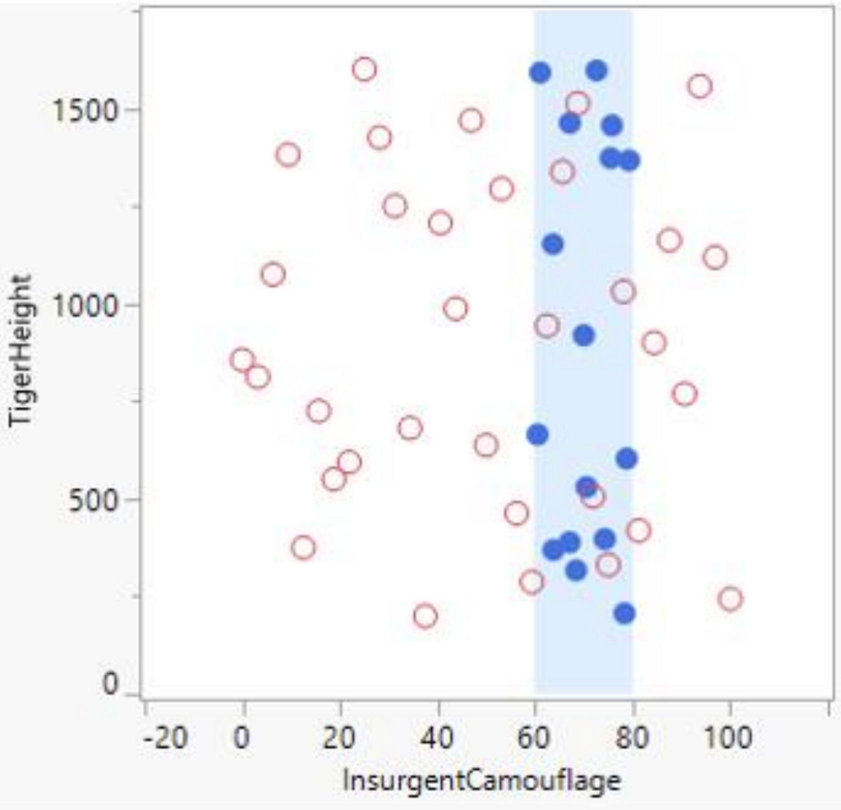
PropIdentINS vs. *Camouflage* at Different *Heights*

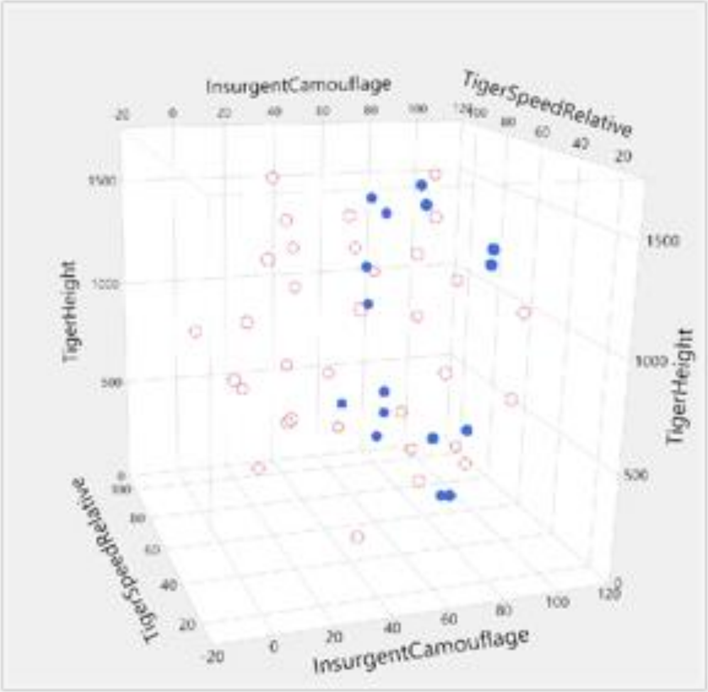




jmp STATISTICAL DISCOVERY FROM SAS

Possible to design Trials Sequentially (33 + 16) to focus attention on ***Insurgent Camouflage*** between 60 and 80





Use Fast Flexible Filling (cluster of random points) with MaxPro criterion.

Go to JMP

- Create 33-trial LHC Design
  - Augment with 32 FFF trials over full range
  - Augment with 16 FFF trials in narrow range for InsurgentCamouflage

Honest Assessment

Create Stratified and Grouped Validation Column by Breaking Data into *Training* +, *Validation* (Tune) X, & *Test* □ Subsets

(Used in model selection & estimating prediction error on new data.)

Stratification Columns: PropIdentINS
Grouping Columns: Excursion

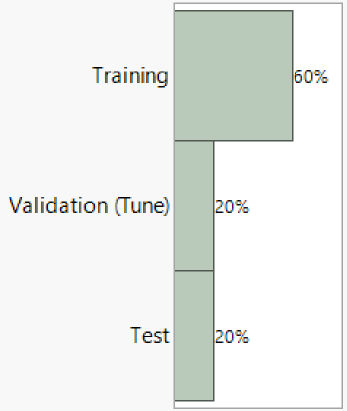
Specify rates or relative rates

		Adjusted Rates	Group Counts
Training Set	0.6	0.6	39
Validation Set	0.2	0.2	13
Test Set	0.2	0.2	13
Excluded Groups			0
Total Groups			65

Options

New Column Name: Strat Group Val 39-13-13
Validation Column Type: Fixed
Random Seed: 12345

Strat Group Val 39-13-13

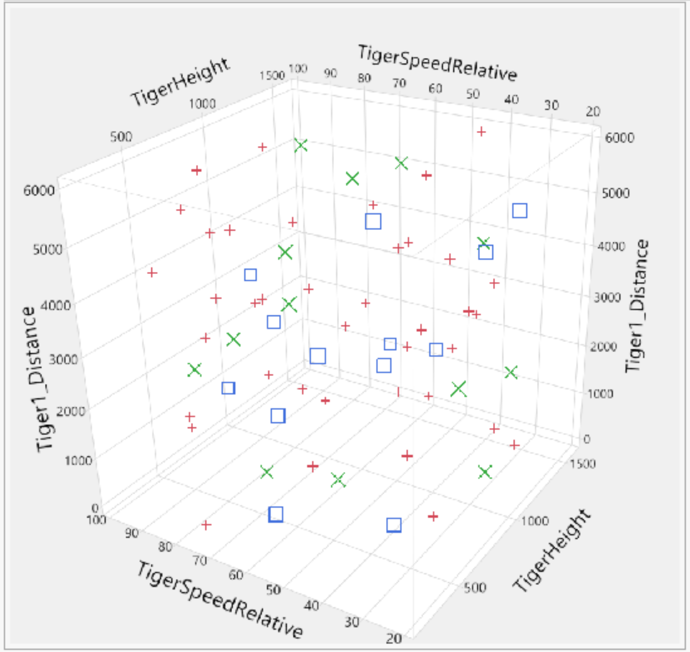


Frequencies

Level	Count	Prob
Test	1296	0.20068
Validation (Tune)	1286	0.19913
Training	3876	0.60019
Total	6458	1.00000

N Missing 0
3 Levels

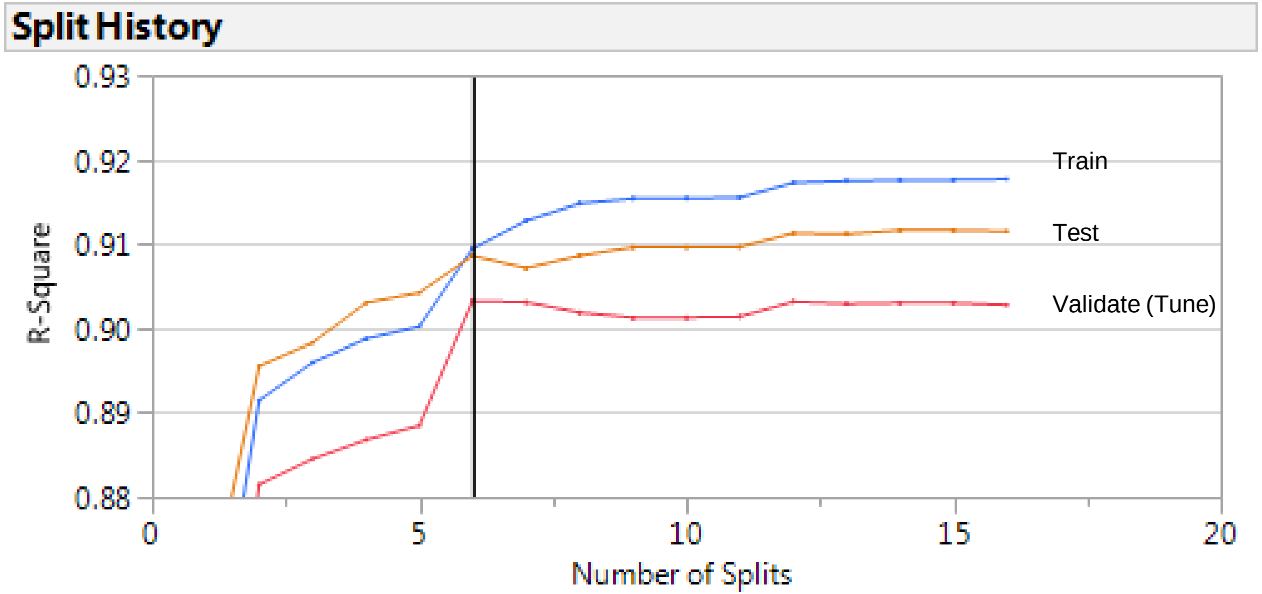
Scatterplot 3D



The Elements of Statistical Learning – Data Mining, Inference, and Prediction
Hastie, Tibshirani, and Friedman – 2001 (Chapter 7: Model Assessment and Selection)

jmp STATISTICAL DISCOVERY FROM SAS

R-Square vs. Number of Partition Tree Splits for Train, Validate(Tune), & Test data subsets

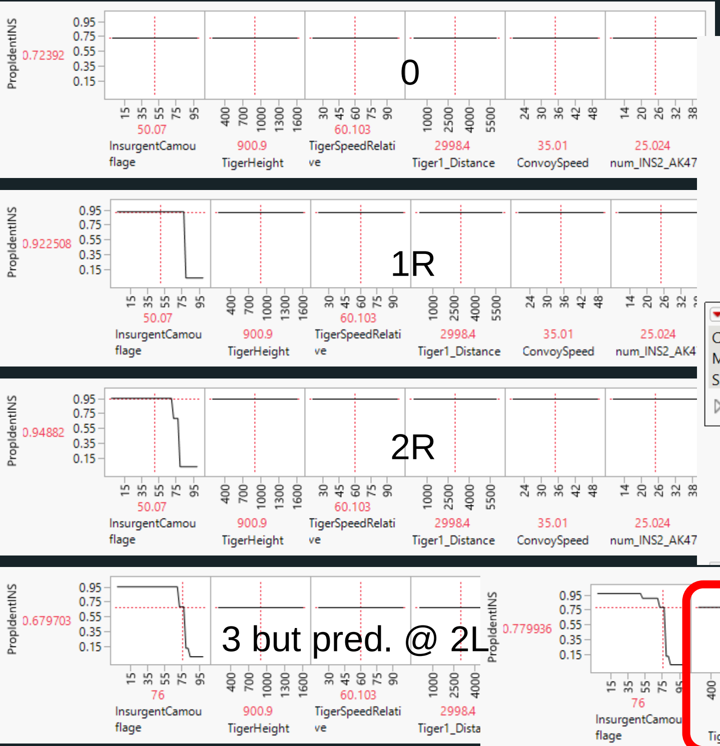


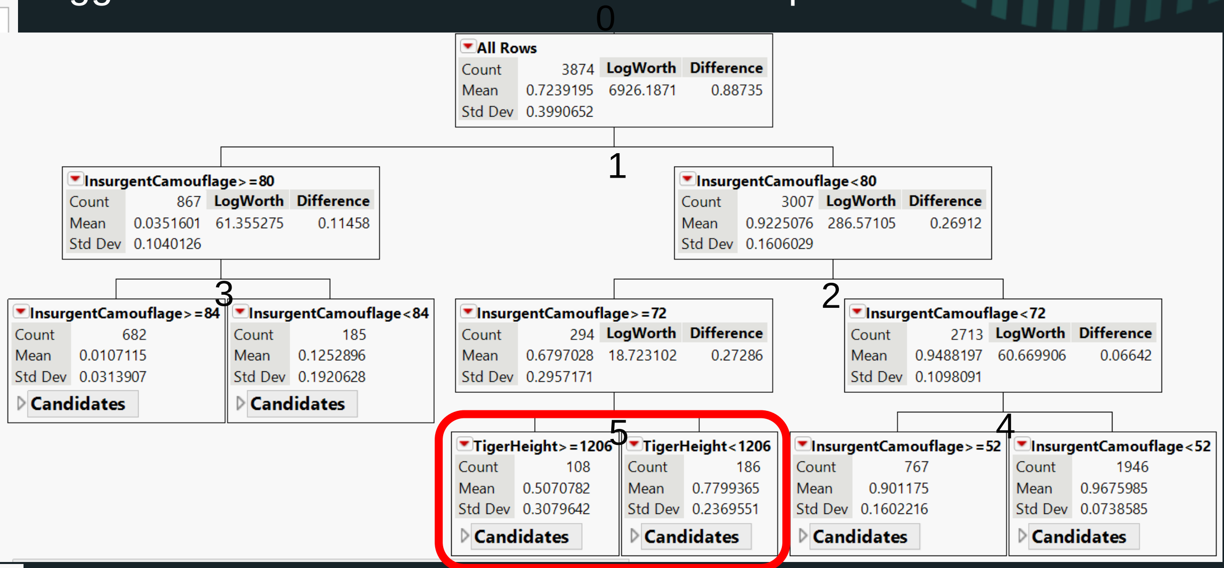
Validation Data in Red
Test Data in Orange

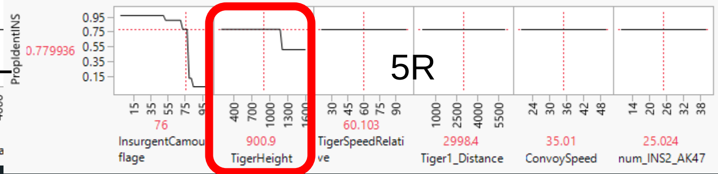
jmp STATISTICAL DISCOVERY FROM SAS

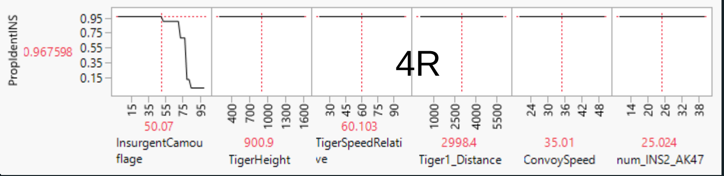
Decision Tree

Each split finds the cut point among all factors that creates the biggest difference in the means of the two partitions of the data

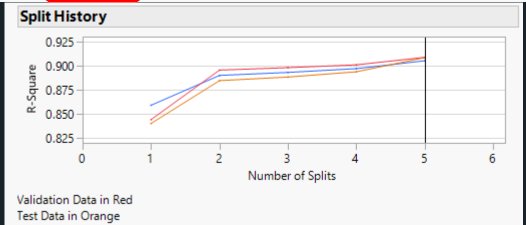








Can be interpreted as a series of nested "If" statements

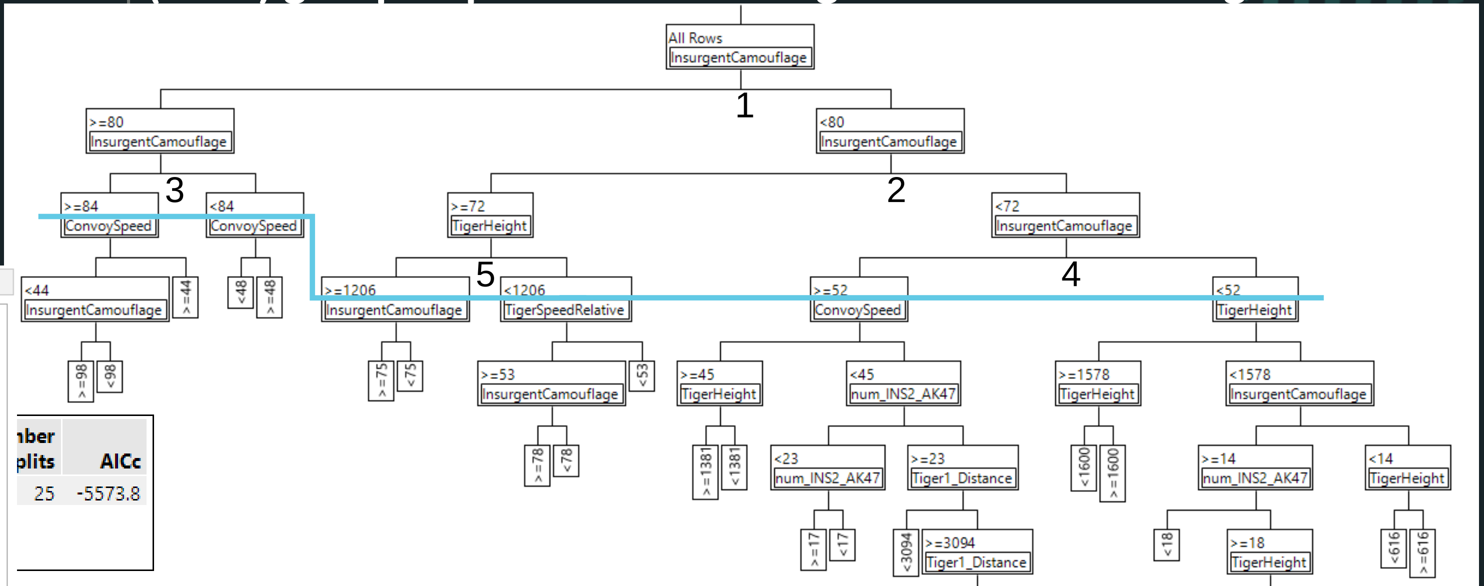


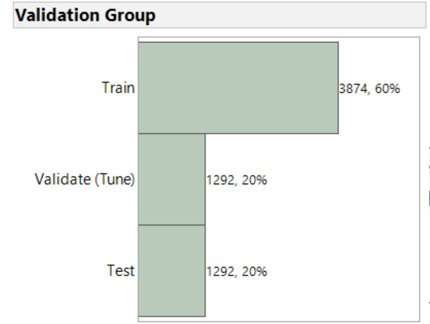
Column Contributions

Term	Number of Splits	SS	Portion
InsurgentCamouflage	4	553.432098	0.9909
TigerHeight	1	5.08702203	0.0091
TigerSpeedRelative	0	0	0.0000
Tiger1_Distance	0	0	0.0000
ConvoySpeed	0	0	0.0000
num_INS2_AK47	0	0	0.0000

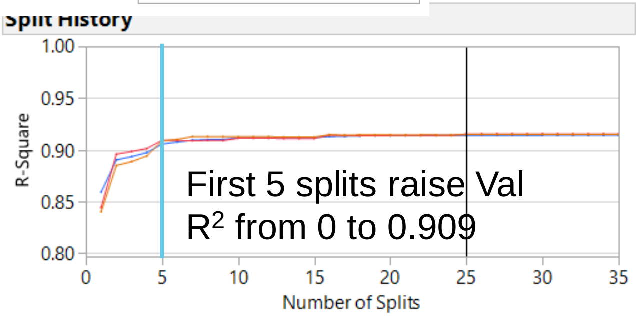
Honest Assessment in Action

Subset data to create *Train*, *Validate*(*Tune*), & *Test* groups. Then use Validate(Tune) group to prevent overfitting of machine learning models.

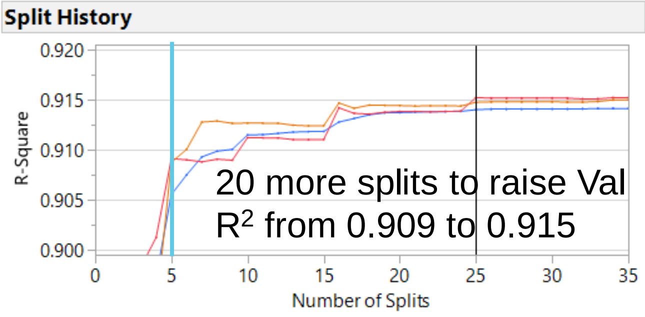




...ber ...plits	AICc
25	-5573.8



Validation Data in Red
Test Data in Orange



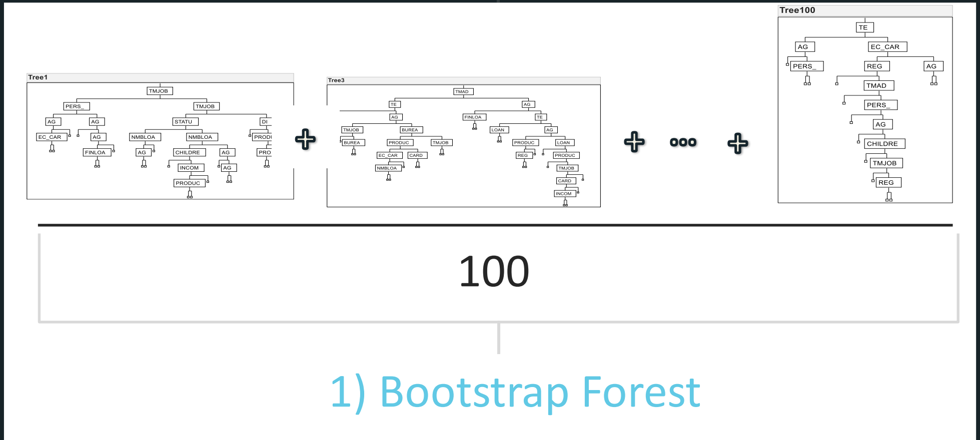
Validation Data in Red
Test Data in Orange

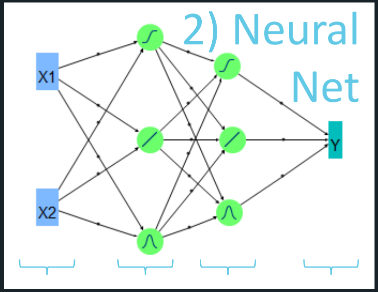
Column Contributions

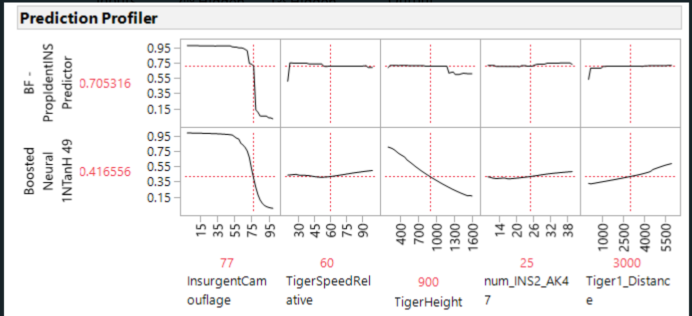
Term	Number of Splits	SS	Portion
InsurgentCamouflage	9	555.084982	0.9847
TigerHeight	6	6.46096421	0.0115
ConvoySpeed	4	1.45893941	0.0026
num_INS2_AK47	4	0.66588349	0.0012
Tiger1_Distance	2	0.06006294	0.0001
TigerSpeedRelative	0	0	0.0000

Robust Strategy for Machine Learning

1) **Bootstrap Forest** – *FAST even w/many Xs – Unlikely to Miss Important Factors*
2) **Neural Network** – *Often Most Flexible & Best Predictor – Tendency to Overfit*
3) **Penalized Regression** – *Often More Interpretable Model + Confidence Intervals*

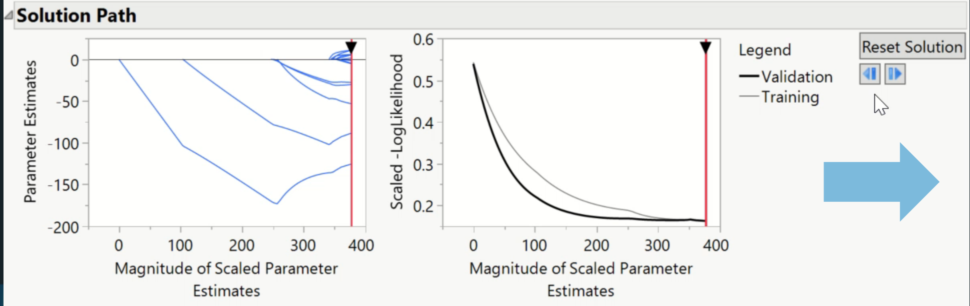


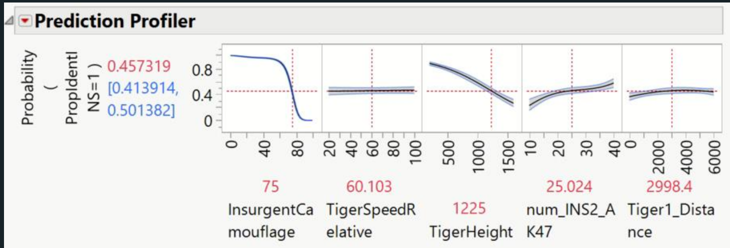




Measure	Training	Validation	Test
Generalized RSquare	0.8026309	0.8019205	0.8069096
Lambda Penalty	0.2258932	.	.

3) Penalized Regression

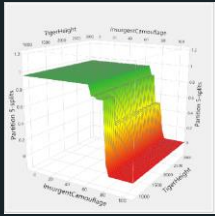


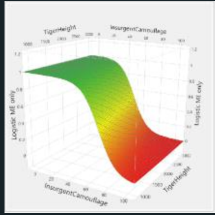


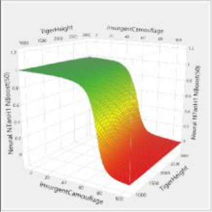
NOT just better prediction, but better understanding!

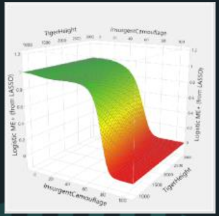
Visually Compare Multiple Machine Learning Models

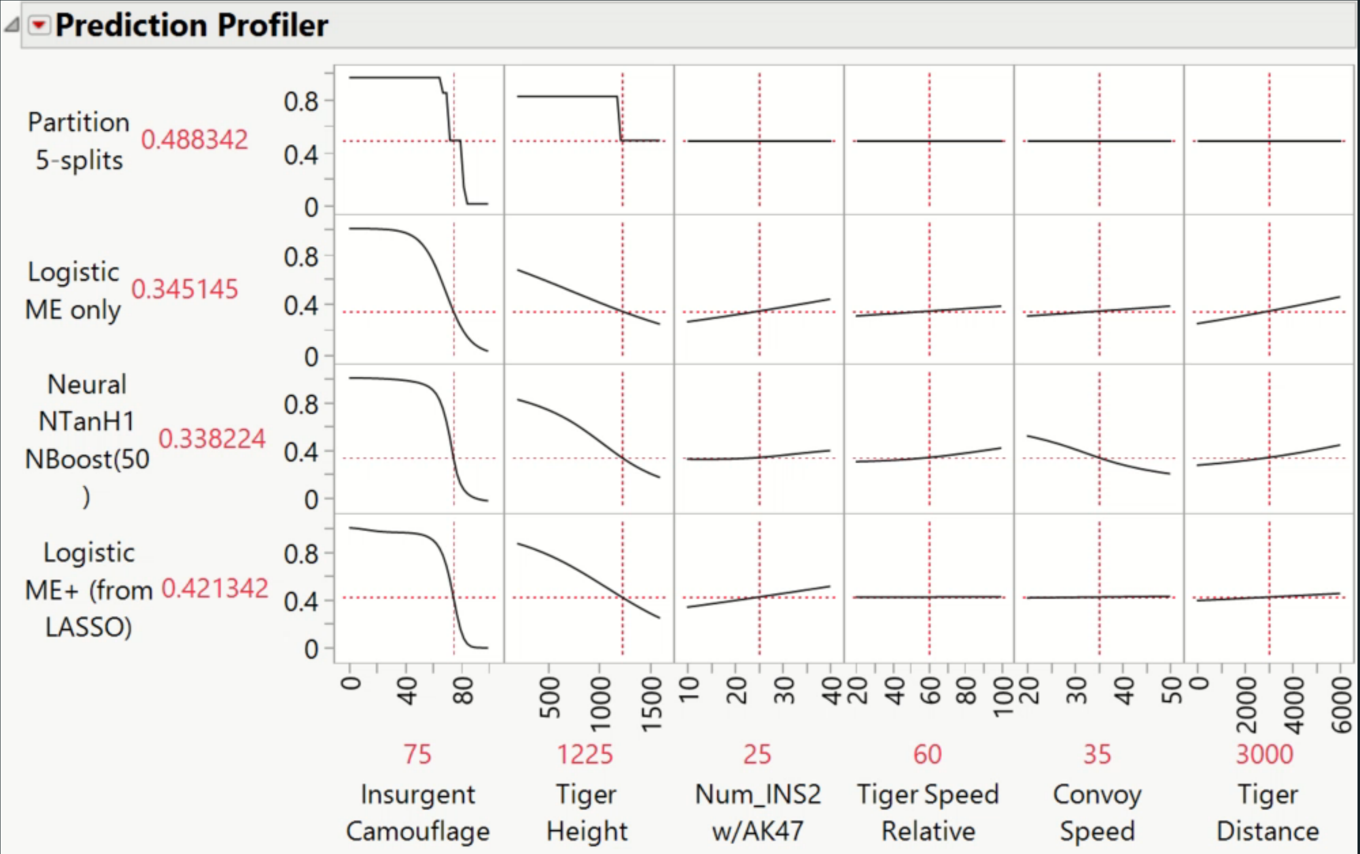
Logistic Regression, Partition with 5-Splits, Neural Network, & LASSO Binomial











Simplest

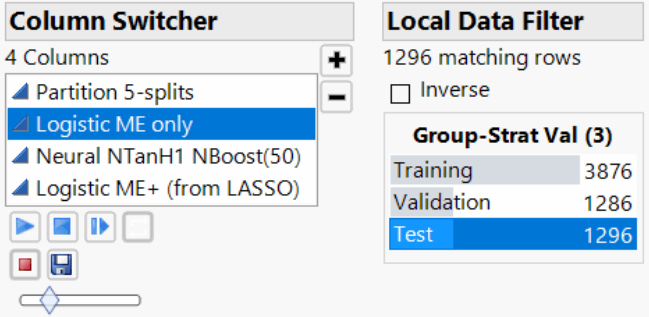
Classic

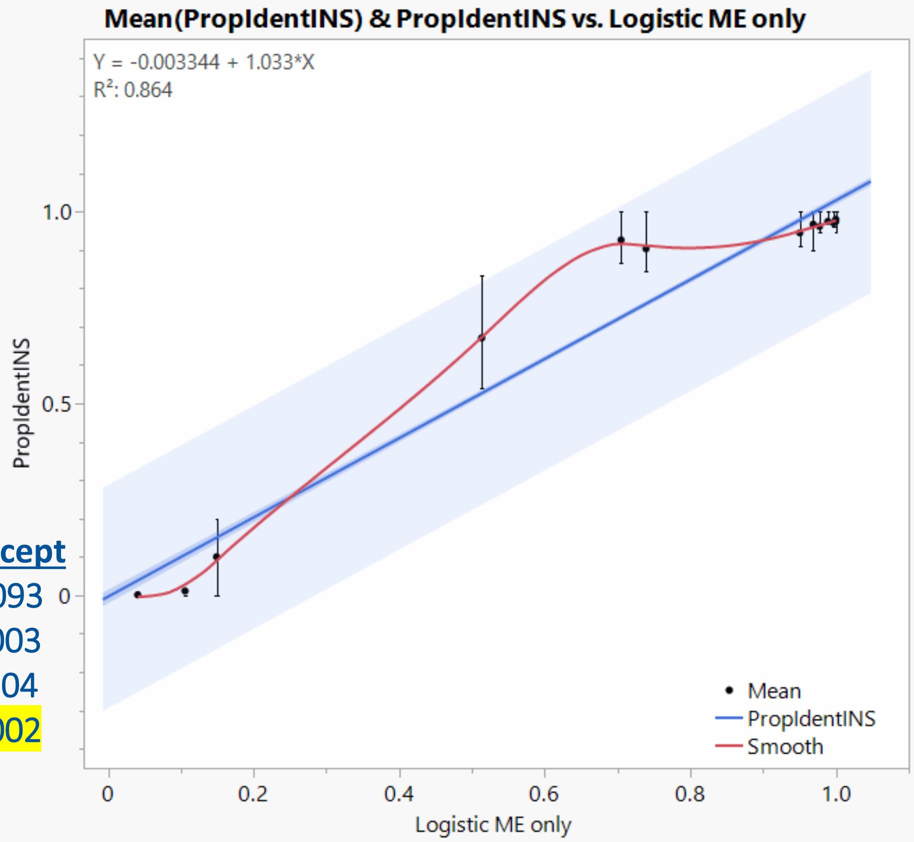
“Best” Predictor

More Interpretable

“LASSO” stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

Actual vs. Prediction Plots for 4 Surrogate Models for *Test Data ONLY* (Not used in fitting or tuning the models)





Where(Group-Strat Val = Test)

Each error bar is constructed using the upper and lower quartiles.

Models	R^2	Slope	Intercept
Partition with 5-Splits	0.847	1.06	-0.093
Logistic Regression	0.864	1.03	-0.003
Neural Network	0.914	0.99	0.004
LASSO Binomial	0.913	1.00	-0.002

Why Use Design of Experiments Methods with Simulation Experiments?

Quicker answers, lower costs, solve bigger problems

- Obtain a ***fast surrogate model*** of the ***slow running simulation***
  - Individual simulations can run for hours, days, weeks
    - Computational Fluid Dynamics (CFD) or Simulation runs in real-time
  - Numbers of factors can be very large (100+)
  - Numbers of simulations needed can be large (thousands in many cases)
  - Simulations can be stochastic requiring many replications
- Surrogate model is a ***fast approximation*** of the simulation
  - more rapidly answer "what if?" questions – ***Instantaneous answer for any "NEW" scenario!***
  - do sensitivity analysis of the control factors
  - optimize multiple responses and make trade-offs
- By running sequences of designs one can be as ***cost effective as possible*** & ***run no more trials than are needed*** to get a useful answer
- By running efficient subsets of all possible combinations (DOEs), one can – for the same resources and constraints – ***solve bigger problems***

Takeaways

DOE Part

- JMP provides both *"Traditional"* (2 or 3-level) RSM, Factorial, OA, DSD & GOSSD, AND *Space-Filling* (many level) Latin HyperCube (LHC), Fast Flexible Filling (FFF), and more designs for computer experimentation
- After running and analyzing a DOE for a computer simulation, one can provide an *instantaneous answer for any new scenario*

Surrogate Modeling Part

- Able to deliver an *interactive trade-space analysis*
- Honest assessment approach of splitting data into TVT subsets makes applying machine learning methods *robust to overfitting*
- *Robust Machine Learning Strategy*
  - 1) *Bootstrap Forest* – fast even with many Xs – unlikely to miss factors
  - 2) *Neural Network* – often most flexible and best predictor
  - 3) *Penalized Regression* – often more interpretable + confidence intervals

Questions?

Thanks for watching today's Mastering JMP.

https://community.jmp.com/

Discussions
Solve problems and share tips & tricks with other JMP users.

File Exchange
Download and share JMP add-ins, scripts, and sample data.

JMP Blogs
Read about a broad range of data analysis topics and posts that inform your JMP use.

Learn JMP
Extend your JMP skills with on-demand videos and JMP files.

JMP Wish List
We want to hear your ideas for improving JMP. Share them here.

JSL Cookbook
Building blocks of JSL code to reduce your coding workload.

JMP Users Groups
Meet up and discuss with other JMP users near you.

Discovery Summit
Info on upcoming Summits and materials from past events.

Community Help
Help with getting started, finding things, and how the Community works.

***Tom Donnelly**, PhD, CAP*
Principal Systems Engineer
JMP Defense & Aerospace Team
Tom.Donnelly@jmp.com
302-489-9291

Additional Slides from Full Tutorial

JMP STATISTICAL DISCOVERY FROM SAS

All 648 Possible Combinations of Settings for 6 Variables (6 X 2 X 2 X 3 X 3 X 3)

TBM = 1,

HoB = 0

TBM = 1,

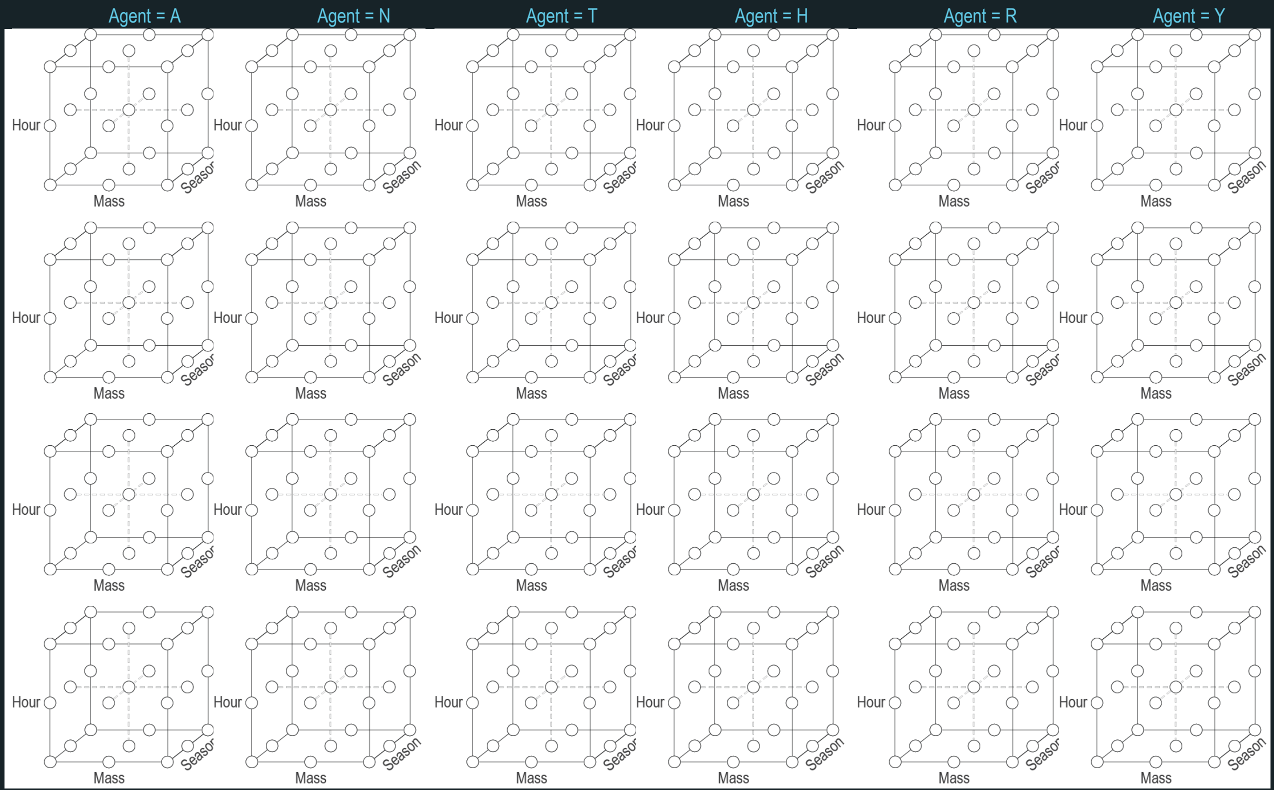
HoB = 10

TBM = 2,

HoB = 0

TBM = 2,

Hob = 10



TBM = 2,

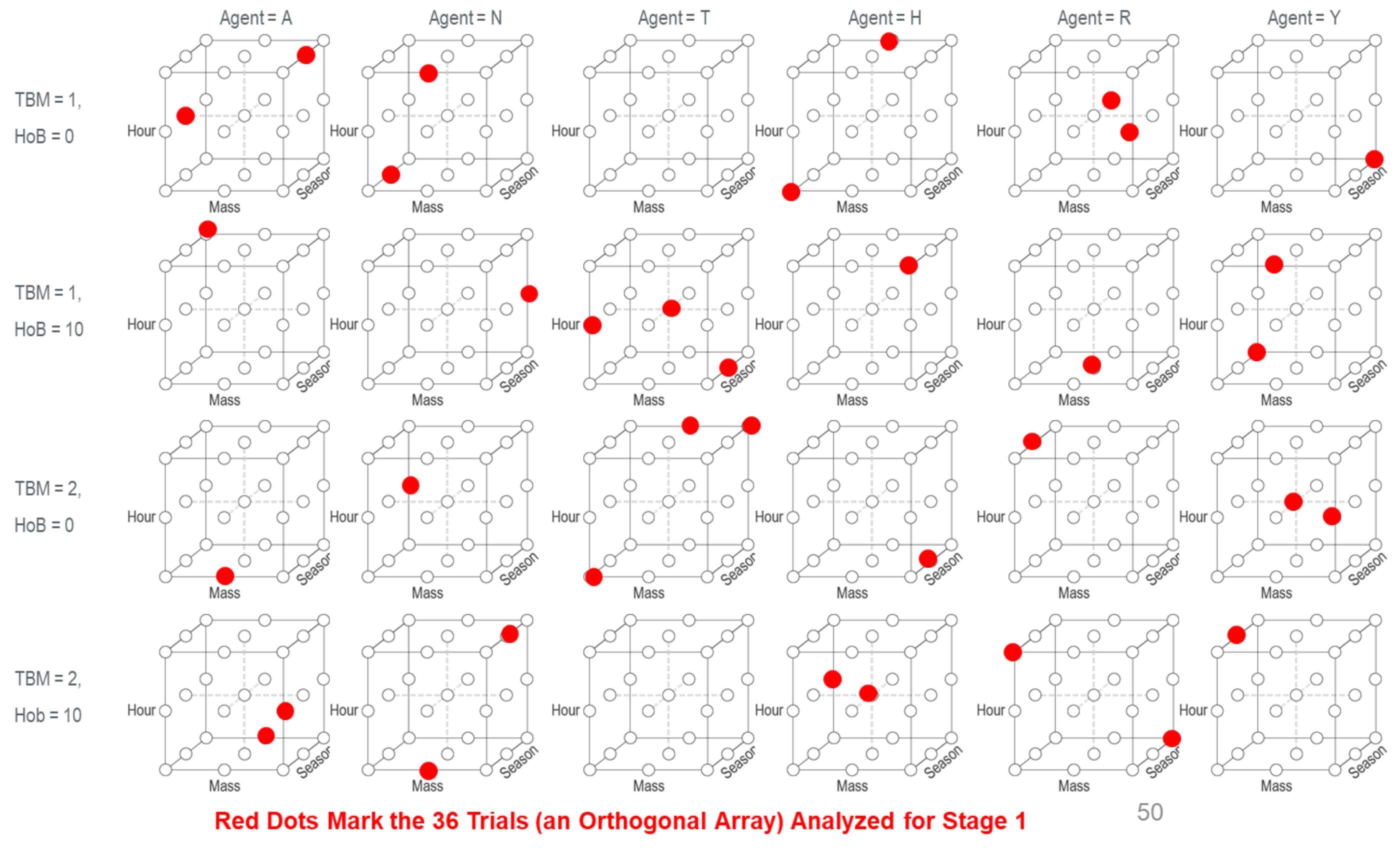
Hob = 10

Four Stage Design Sequence

Stage 1	Stage 2	Stage 3	Stage 4
36 Total Simulations	108 Total Simulations	324 Total Simulations	ALL 648 Simulations
Design 1, 36 trials	Design 1, 36 trials	Design 1, 36 trials	Design 1, 36 trials
	Design 2, 72 trials	Design 2, 72 trials	Design 2, 72 trials
		Design 3, 216 trials	Design 3, 216 trials
Main effects only for ALL variables + some 2-way interactions	Stage 1 effects plus all 2-way interactions + some 3-way interactions	Stage 2 effects plus all 3-way interactions	Stage 3 effects plus ALL remaining 4-way, 5-way and 6-way interactions
5.6% of 648	16.7% of 648	50% of 648	**Design 4, 324 trials** NOTE: Length of this green box should be longer than shown

324 trials in Design 4 used as checkpoints for Designs 1, 2 & 3 →

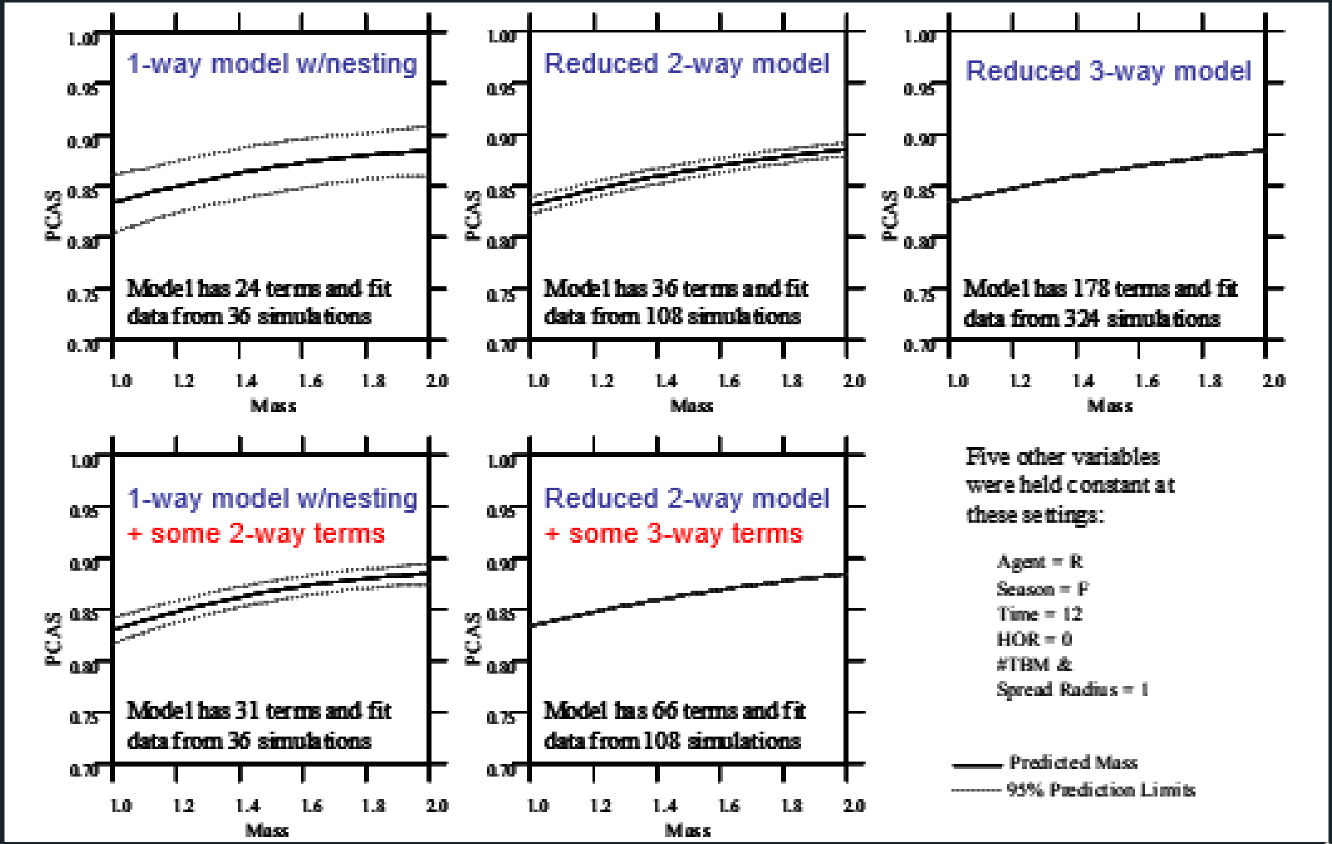
36 of All 648 Possible Combinations of Settings for 6 Variables (6 X 2 X 2 X 3 X 3 X 3)



Can Get Designs from Different Sources

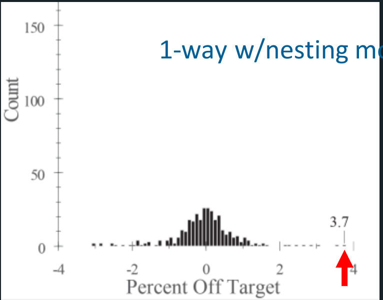
- Textbook
  - Limited number of catalogued solutions – experimenters frequently change their problem to match available designs
  - Variable settings are in coded units
- Web sites of designs
  - Greater number of catalogued solutions – but never all
  - Variable settings are in coded units
- Custom computer code
  - Can find solutions for previously un-catalogued cases
  - Variable settings are in coded units (-1, 0, 1)
- COTS Solution
  - Textbook and algorithmic code for generating custom designs
  - Variable settings in natural or laboratory units (120, 150, 180)

Predictions (w/ 95% Pred. Limits) of PCAS vs. Nested Mass and MunCnt_Spread for 1-way, reduced 2-way and reduced 3-way models



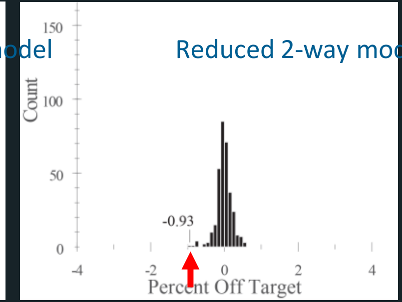
“Factor Sparsity” and “Effect Heredity” Used to Enhance Model Complexity

36 trials



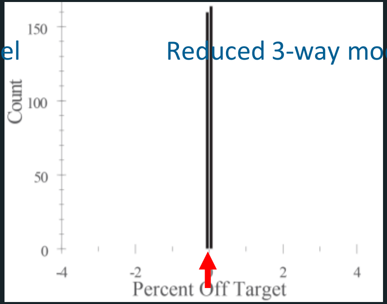
Worst Case = 3.7%
Half of Cases < 0.37%

108 trials

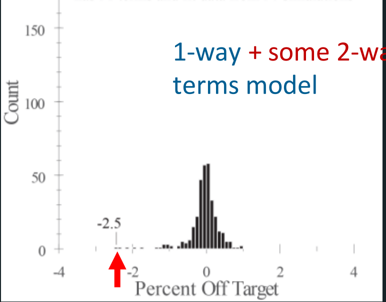


Worst Case = -0.93%
Half of Cases < 0.11%

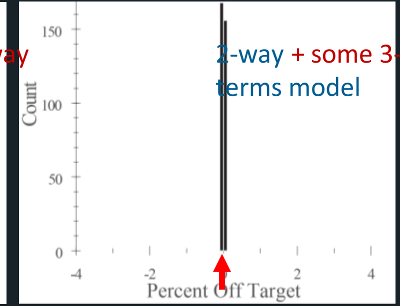
324 trials



Worst Case = -0.0081%
Half of Cases < 0.0007%



Worst Case = -2.5%
Half of Cases < 0.16%



Worst Case = -0.0251%
Half of Cases < 0.0010%

Factor Sparsity states only a few variables will be active in a factorial DOE

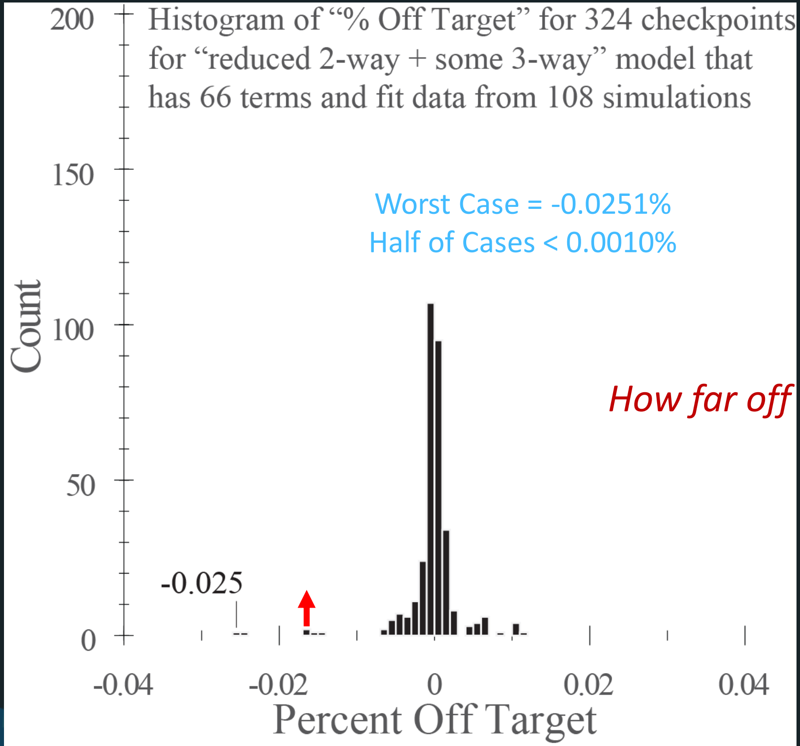
Effect Heredity states significant interactions will only occur if at least one parent is active
(See Wu & Hamada, p. 112)

Oct. 1, 2007 visit by Profs. Wu & Joseph of GA Tech ISyE

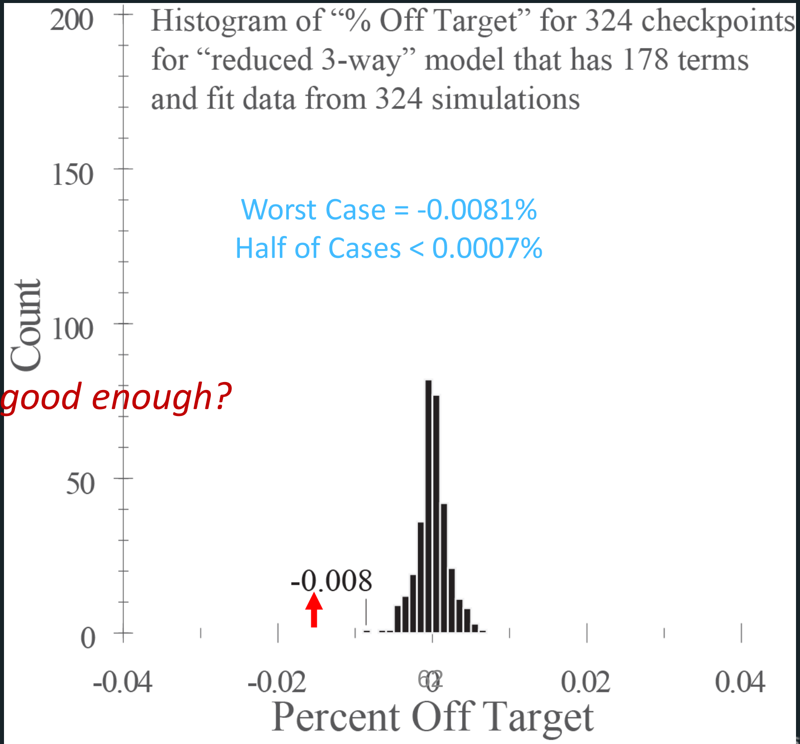
Only a Fraction of All Possible Trials May be Required to Provide an Answer

108 trials **324 trials**

Higher Resolution (100X) Histograms of the "Percent Off Target" that Response Predictions Fell Relative to 324 Checkpoint Observations



How far off is good enough?



Conclusions for Sequential Traditional Designs

- Possible to get the 80% to 95% solution with less than 20% of the brute force running of all factor combinations
- Use of “factor sparsity” and “effect heredity” principles can help to get more information than the design was originally built to support
- Next stage trials can first be used as checkpoints for previous stages
- With improved efficiency over running all combinations, more factors can be studied with the same resources

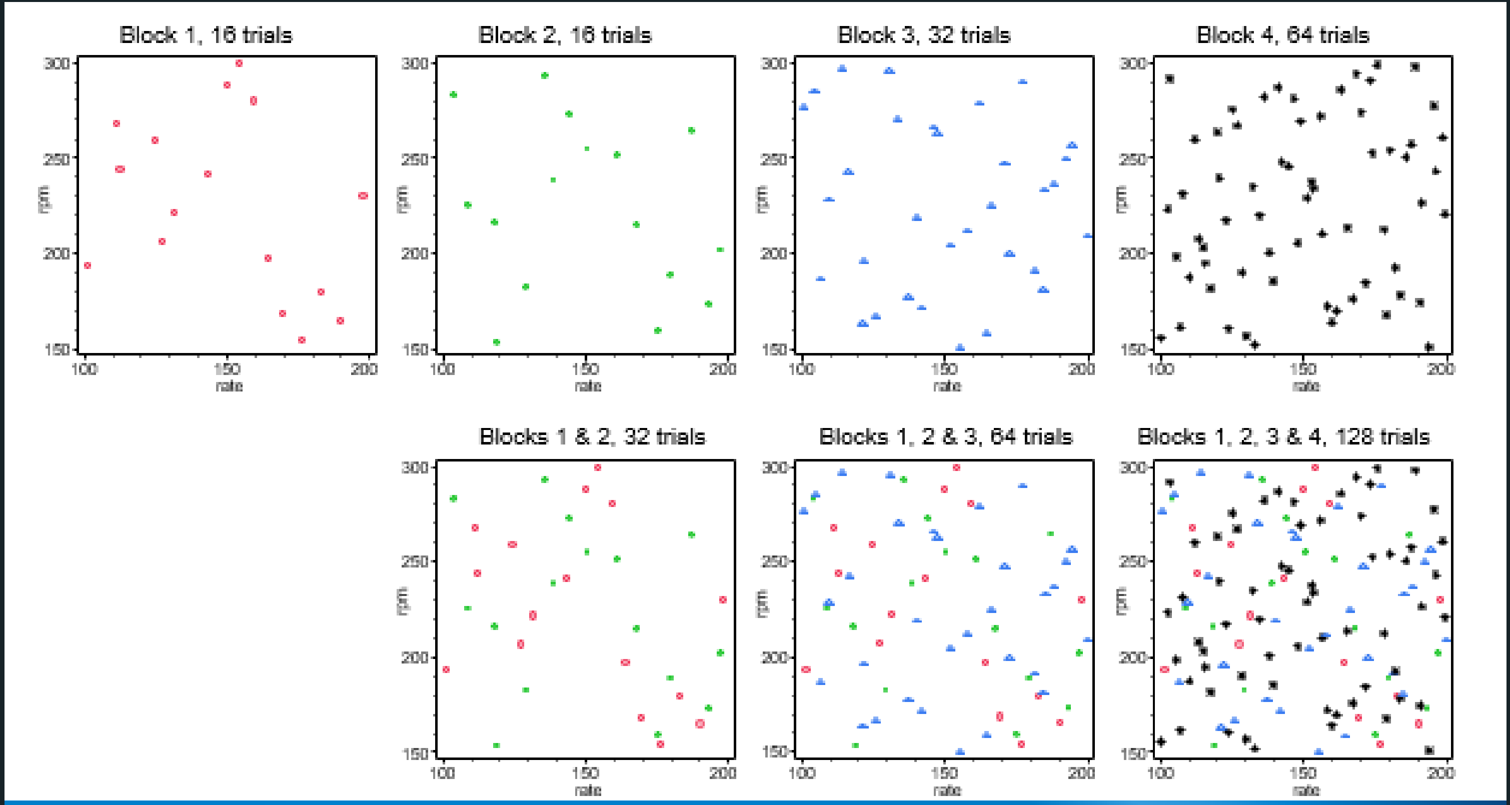
Why Is a Sequential Approach So Useful?

We wanted to not just do sensitivity analysis of the factors, but **provide an interactive surrogate model of the long-running simulation so that analysts could evaluate "what if?" scenarios.**

The problem was that the Computational Fluid Dynamics models we were looking to run could take a week on a single CPU or **12 hours on 64 CPU cluster**. With on the order of 10 factors we expected to need to run on the order of **100 simulations**. **This meant it could be weeks or months before we could start our analysis.**

Nested Latin Hypercube Designs gave us a way to start analyzing data after about the first 20% of the simulations were run. We also wanted to be able to run just enough simulations to achieve a surrogate model accuracy of 90%.

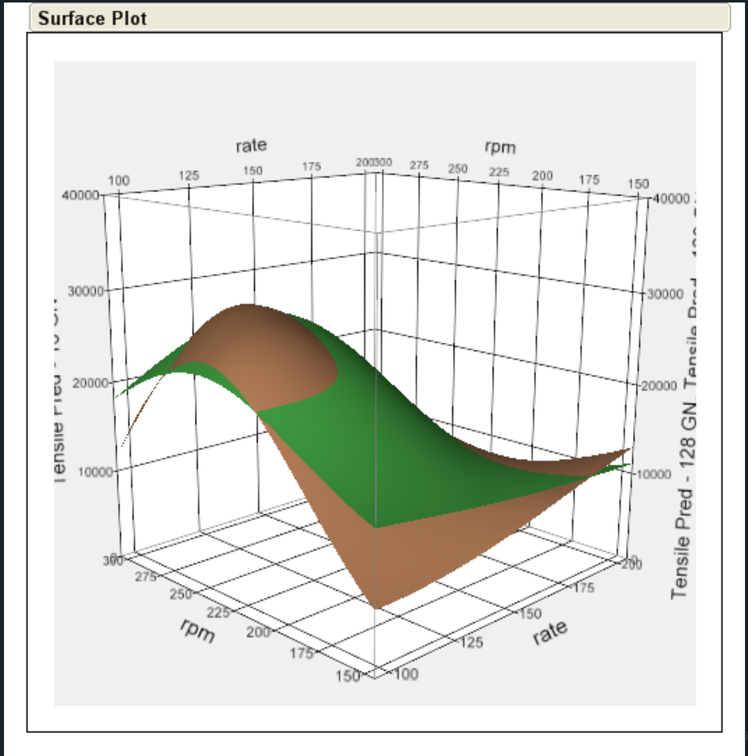
Projections of Trial Locations in 2 factors for a 10-factor, 128-trial, Nested Latin Hypercube Design* (NLHD) with 4 Blocks

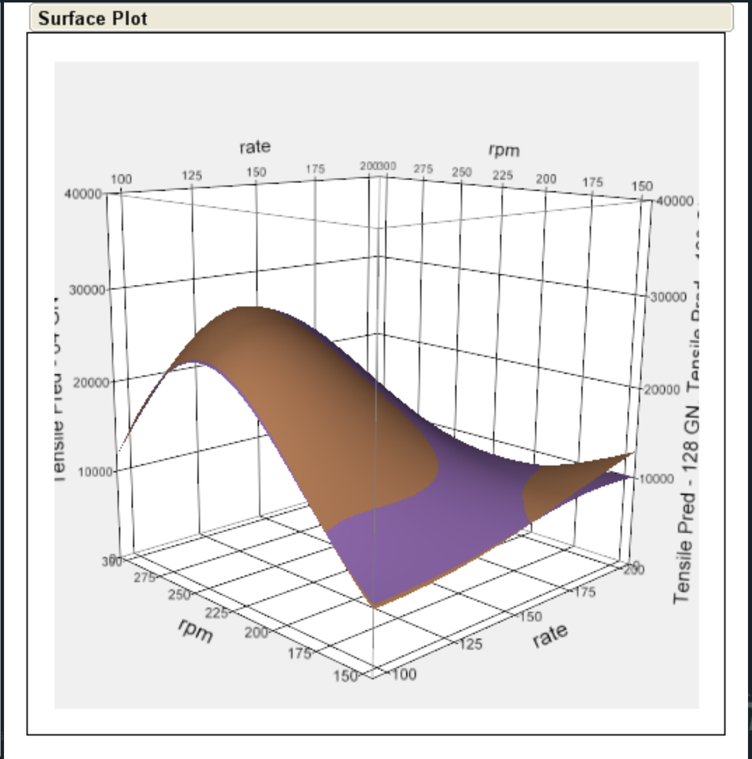


*Generated with Matlab Code Received from Prof. Peter Qian of U of Wi.

Compare Response Surfaces for fit of 16 vs. fit of 128 trials (left) and for fit of 64 vs. fit of 128 trials (right)

Stage 1 fit of16 trials colored green
Stage 4 fit 128 trials colored brown
Stage 3 fit 64 trials colored purple

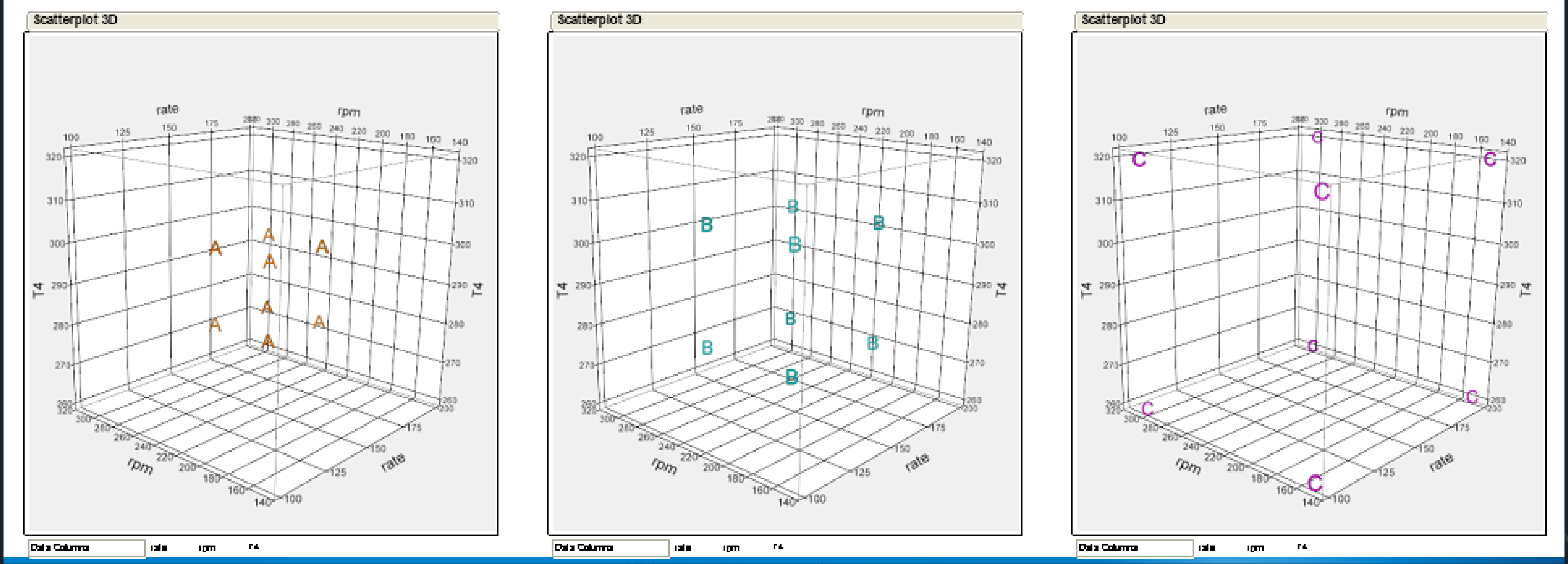




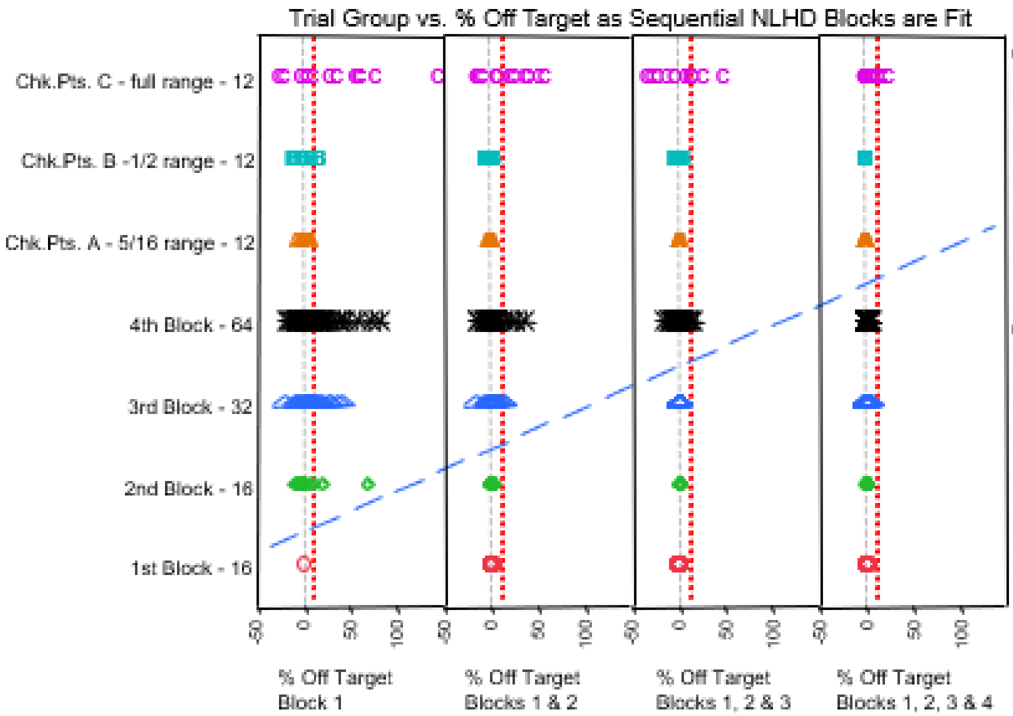
Why Run Simulations in Sequential Blocks?

The point of running this sequence of blocks is to be able to evaluate the surrogate model after each stage to see how accurately it is predicting observed values of 3 sets of checkpoint trials. If it proves to be sufficiently accurate, then subsequent blocks of simulation trials need not be run.

Without the NLHD approach one has to choose the "right" size space-filling design in order to get useful results. If you choose too small a design, one has to start over with a larger design.



Accuracy of Surrogate Predictions for 3 Groups of Check-points Yielding Marginal, Moderate and Extreme Extrapolation

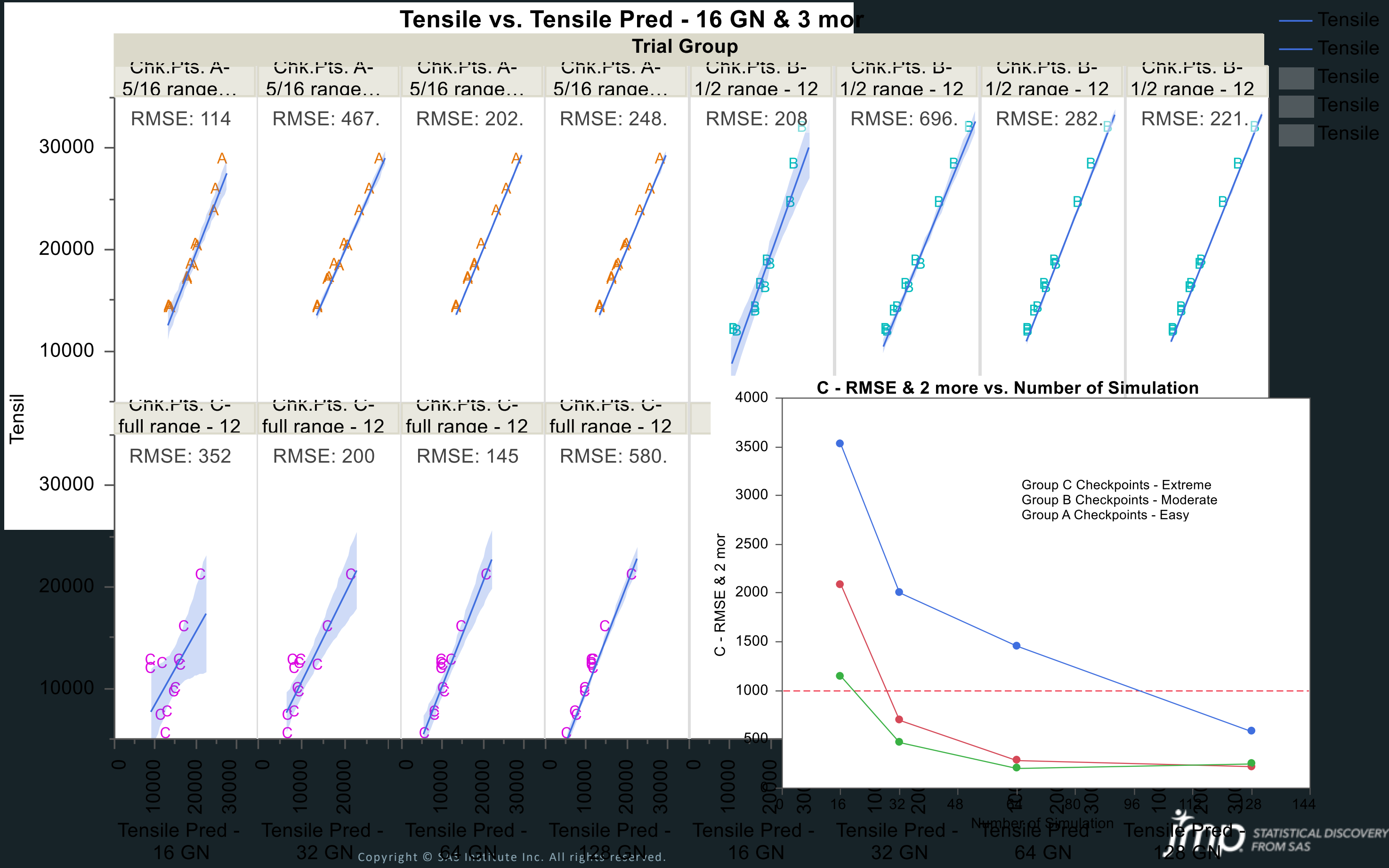


% Off Target for Checkpoints NOT Included in Model Fit

% Off Target for Points Fit with Gaussian Process Modeling

Plots of Actual vs. Predicted (Simulation vs. Surrogate) by Checkpoint Group for 4 Stages of Analysis of NLHD

Checkpoint Groups A & B show diminishing return in prediction improvement for running past stage 3



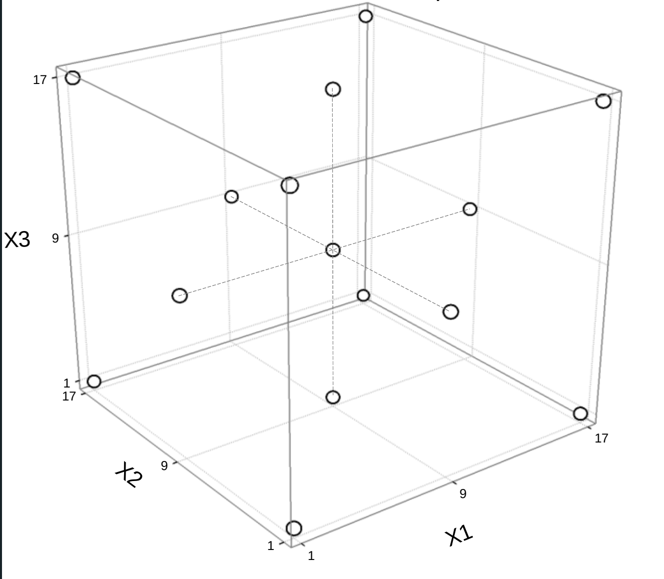
jmp STATISTICAL DISCOVERY FROM SAS

Conclusions Sequential Space-Filling Designs

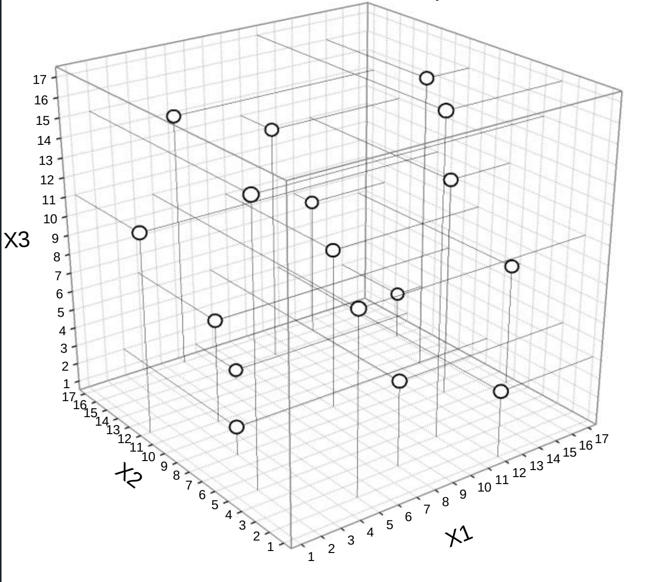
- NLHD designs can be run sequentially so that surrogate model accuracy can be evaluated after each block and decision made as to whether or not to move forward with the next block
- Generally, as more NLHD blocks are run, the surrogate model accuracy increases
- Inclusion of extreme (full range) extrapolation checkpoints will expand interpolation volume of Kriging analysis – assuming Kriging analysis remains stable
- Caveat: These conclusions were reached using a moderately complex transcendental function in lieu of a CFD simulation model that is believed to do a good job of stressing extrapolation with the surrogate model..

How are Space-Filling Designs Different from Traditional Designs?

Response-Surface Design
for 3-Variables with 15 Unique Trials



Space-Filling Design
for 3 Variables with 17 Unique Trials



Rather than emphasizing high leverage trials (“corners”) for a simple polynomial model, space-filling designs “spread” their trials more uniformly through the space to better capture the local complexities of the simulation model.

Websites for Designs, Software & Publications

- The Simulation Experiments & Efficient Design (SEED) Center for Data Farming at Naval Postgraduate School https://nps.edu/web/seed/
  - Designs https://nps.edu/web/seed/software-downloads
    - Nearly Orthogonal Latin Hypercubes (NOLH), nearly orthogonal & balanced (NOB) designs for mixed factor types, 2nd order NOLH, and nearly saturated NOLH (S-NOLH)
    - Resolution V, Fractional Factorials for many factors
  - Agent-Based Simulation Software
    - Pythagoras
    - MANA (Map Aware Non-uniform Automata)
  - Many Papers for Download and Links to INFORMS and WSC
- Library of Orthogonal Arrays maintained by Neil J. A. Sloane
  - http://neilsloane.com/oadir/
- Library of Orthogonal Arrays maintained by Warren F. Kuhfield
  - https://support.sas.com/techsup/technote/ts723b.pdf