Statistical Discovery.™ From SAS.

EFFICIENT MODELING & SIMULATION USING DESIGN OF EXPERIMENTS

22nd NDIA Systems & Mission Engineering Conference
October 24, 2019

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

OUTLINE

- Background & Resources
- Why Use DOE for M&S?
- Why is DOE important?
- Overview of Design of Experiments (DOE)
- Efficient M&S Using DOE – 3 Examples
  - Sequential traditional DOE
  - Space-Filling DOE Case Study
  - Sequential space-filling DOE

USING DESIGN OF EXPERIMENTS (DOE) FOR 35 YEARS

- ‘83-’87 Honeywell, Inc., Engineer
  First saw the power of DOE in 1984
- ‘87-’99 ECHIP, Inc., Partner & Technical Director
  200+ DOE courses, on-site at 40+ companies
- ‘99-’05 Peak Process, LLC, Consultant
- **‘05-’08 US Army, Edgewood Chemical Biological Center (ECBC), Modeling, Simulation, & Analysis Branch**
  DOE with Real data and Modeling & Simulation data
- ’08-’19 SAS Institute Inc., JMP Division
  Data Visualization, Data Analytics, and their synergy with DOE
  Support DoD sites, NASA, & Defense Contractors

PROJECTS USING DOE AT U.S. ARMY ECBC CY05-08

Detection, Decontamination & Protection

- JPM Nuclear Biological Chemical Contamination Avoidance (NBCCA) - Whole Systems Live Agent Test (WSLAT) Team support to the Joint Biological Point Detection System (JBPDS)
- Agent Fate wind tunnel experiments
- Decontamination Sciences Team
  - Contact Hazard Residual Hazard Efficacy Agent T&E Integrated Variable Environment (CREATIVE) - real and simulation data
  - Modified vaporous hydrogen peroxide (mVHP) decontamination – real data
- Smoke and Target Defeat Team
  - Pepper spray characterization – real data
  - Obscurant material evaluation (with OptiMetrics, Inc.) – simulation data
- U.S. Army Independent Laboratory In-house Research (ILIR) on novel DOE used with simulations
  - **Re-analysis of USAF Kunsan AFB Focused Effort BWA simulation data**
  - CB Sim Suite used for sensitivity analysis of atmospheric stability
- U.S. Marine Corps Expeditionary Biological Detection (EBD) Advanced Technology Demonstration (ATD)
  - Chamber testing of detectors – real data
  - CB Sim Suite sensor deployment studies – simulation data
- U.S. Navy lead on Joint Expeditionary Collective Protection (JECP)
  - Swatch and chamber testing – real data
  - **Computational Fluid Dynamics (CFD) – simulation data**

DOWNLOADS

- PDFs available
  - White Paper 2008 - **Efficient Modeling & Simulation of Biological Warfare Using Innovative Design of Experiments Methods** – Tom Donnelly
    https://www.jmp.com/en_us/whitepapers/jmp/modeling-biological-warfare.html

  - Dissertation 2017 - **A framework for the optimization of doctrine and systems in Army Air Defense units using predictive models of stochastic computer simulations** – LTC Brian Wade, Technical Director at TRAC MRY
    https://smartech.gatech.edu/handle/1853/58275

RECORDINGS AT WWW.JMP.COM/FEDGOV

These 12 videos primarily cover Design of Experimenst (DOE) topics.

Custom DOE - JMP 13 (not 14) Make the Design Fit Your Problem (Link to Mastering JMP)	Screening Designs Classic FF & PB, and Modern D-Optimal, Supersaturated, DSD, & Alias-Optimal	Compare Designs How to Choose Better Designs on Multiple Criteria
Advanced Custom DOE - JMP 13 (not 14) Augmentation, Broken Design Repair, & Design from a Candidate Set	Definitive Screening Designs (DSD) Creation & Augmentation	Data Transformations Get Rid of L-o-F, Predictions Make Physical Sense (Link to Mastering JMP)
Mixture DOE Efficiently Blending Ingredients to Optimize a Process (Link to Mastering JMP)	Analyzing DSD DOEs Graphical Methods and Fit Definitive Screening Platform	Power Calculation via MC Simulation Binary Responses & Split-Plot Designs
Efficient M&S Using DOE How to Run Fewer Computer Simulations	Exploratory Data and Root Cause Analyses What to Do When You Don't Have a DOE	Covering Arrays - Rapid Fault Detection in Software & Systems

https://community.jmp.com/t5/US-Federal-Government-JMP-Users/Efficient-M-amp-S-Using-DOE-Methods-JMP-14/ta-p/69547

DOWNLOAD & RECORDING

- 16 Factors
- 50,000 unique cases
- Each 1,000 times
- 50 Million Simulations
- Neural Network
  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.

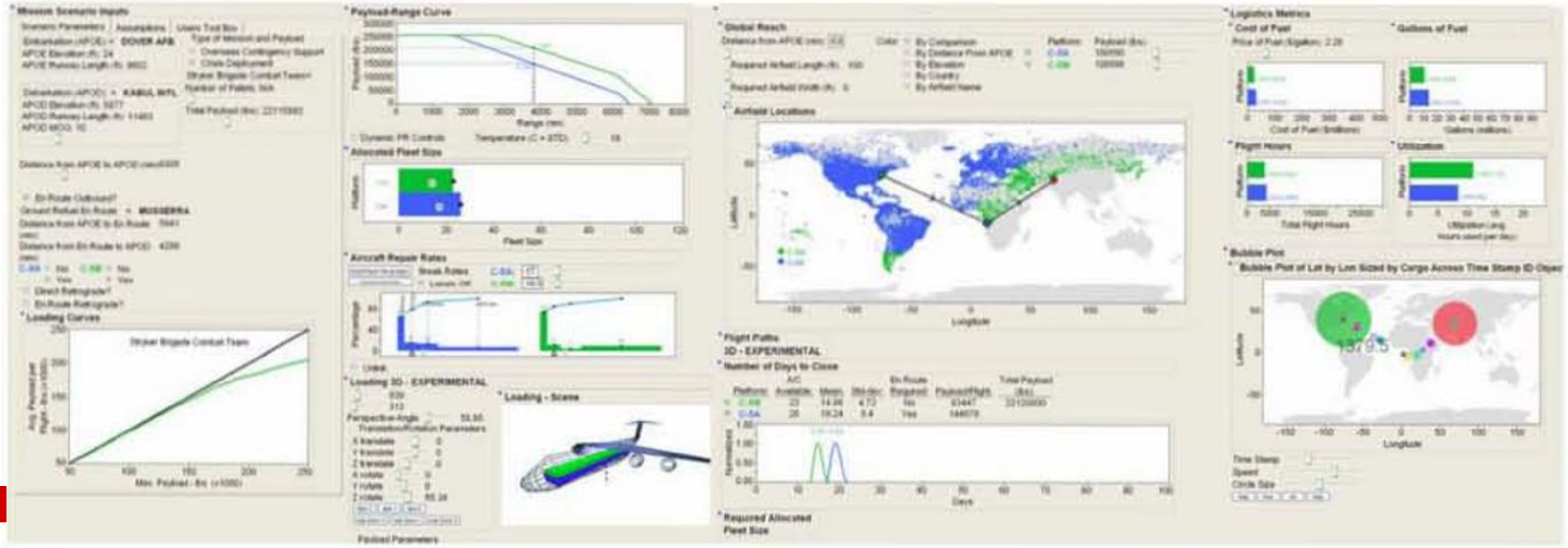


Figure 2. Strategic Airlift Comparison Tool Layout

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

Recording

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

WHY USE DESIGN OF EXPERIMENTS METHODS WITH SIMULATION EXPERIMENTS?

Quicker answers, lower costs, solve bigger problems

- Obtain a fast surrogate model of the 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 yields 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, one can – for the same resources and constraints – ***solve bigger problems***

WHY IS USING DOE IMPORTANT?

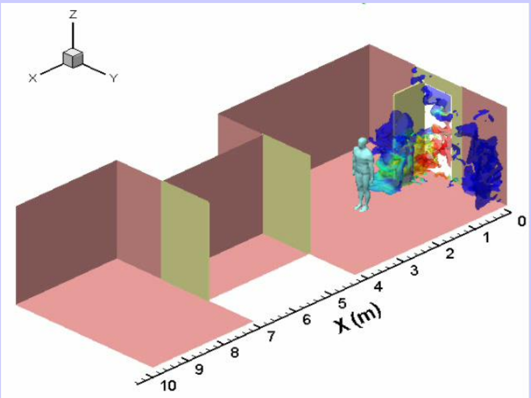
- *“One thing we have known for many months is that the spigot of defense funding opened by 9/11 is closing.”*
- *“In the past, modernization programs have sought a 99 percent solution over a period of years, rather than a 75 percent solution over a period of weeks or months.”*
  - Two quotes from the January 27, 2009 submitted statement of Secretary of Defense Robert M. Gates to the Senate Armed Services Committee.

- DOE is one of the more powerful tools we can use to efficiently accomplish our goals.
  - DOE yields the maximum information from the fewest experiments.
  - DOE often yields an 80% solution in less than 20% of the work.

LONG RUNNING PHYSICS-BASED SIMULATIONS

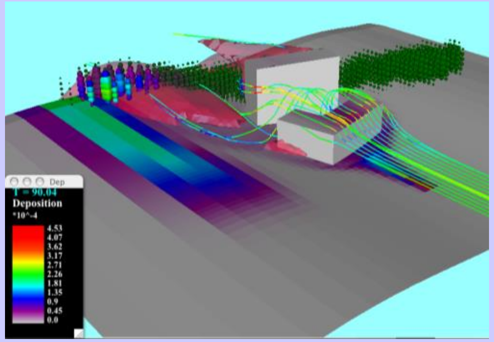
Detailed Physics Models can require a great deal of runtime
to generate a short period of simulation time.

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

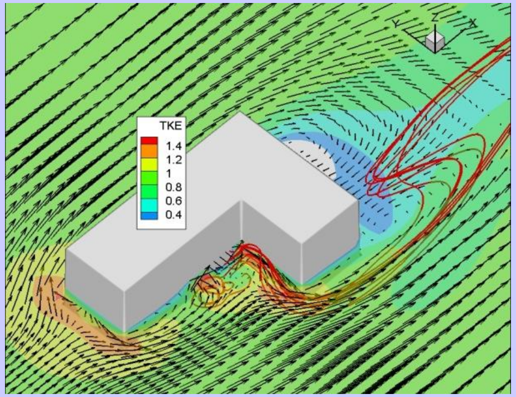
Detailed Ingress/Egress, Internal Airflow and Convection



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

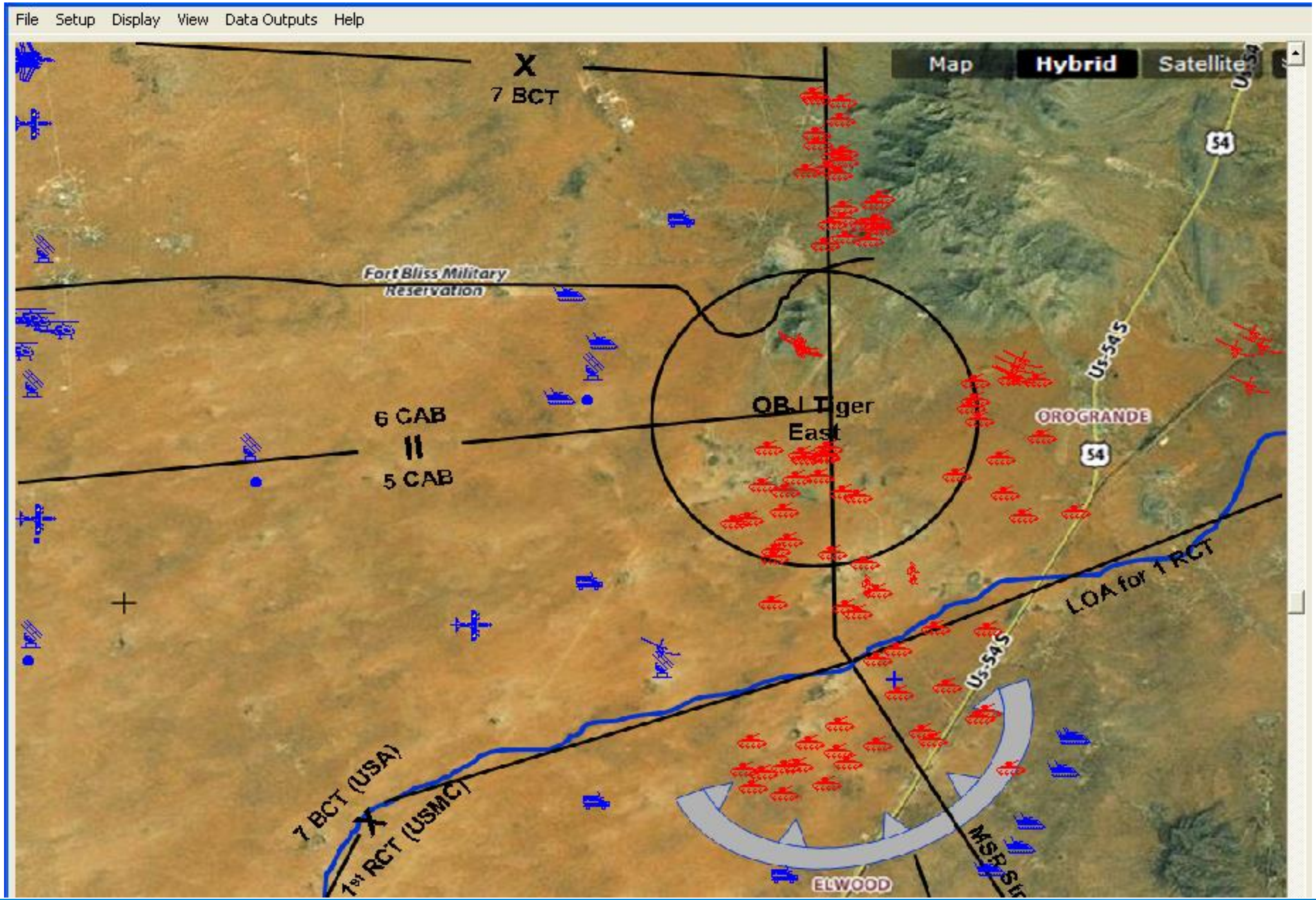


Developed for Exterior
Stationary Grids
TBD Cells
Min-Hours of Simulation
Single CPU
Minutes-Days of Runtime

Speed, Flexibility, More User Friendly, V&V

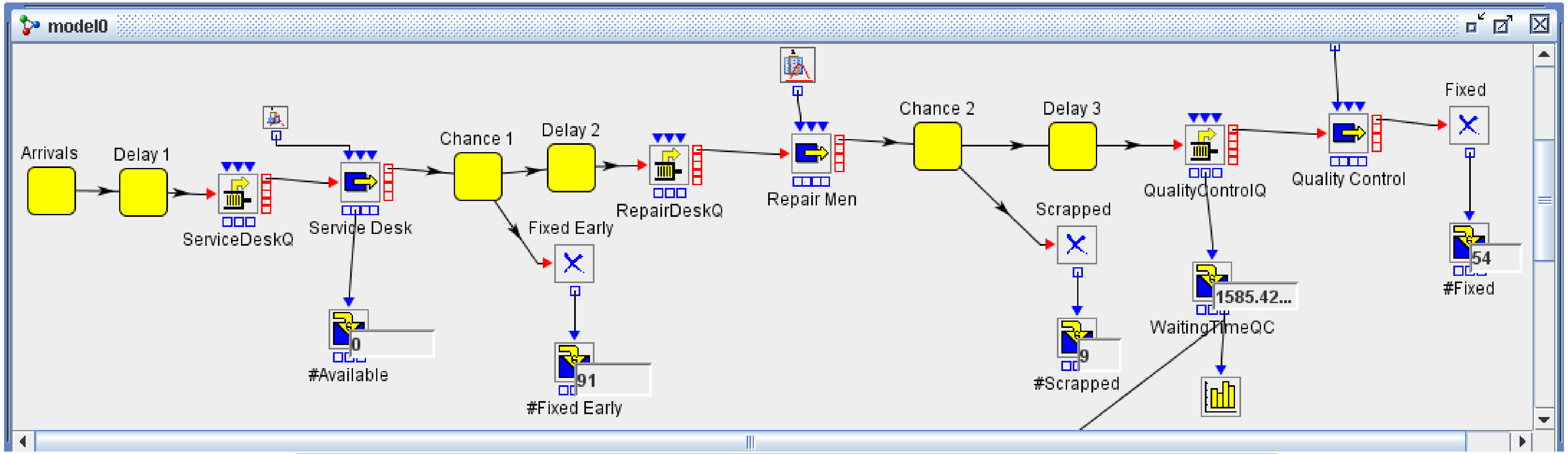
STOCHASTIC SIMULATIONS WITH MANY REPLICATES

Agent Based Simulations



STOCHASTIC SIMULATIONS WITH MANY REPLICATES

Discrete Event Simulations

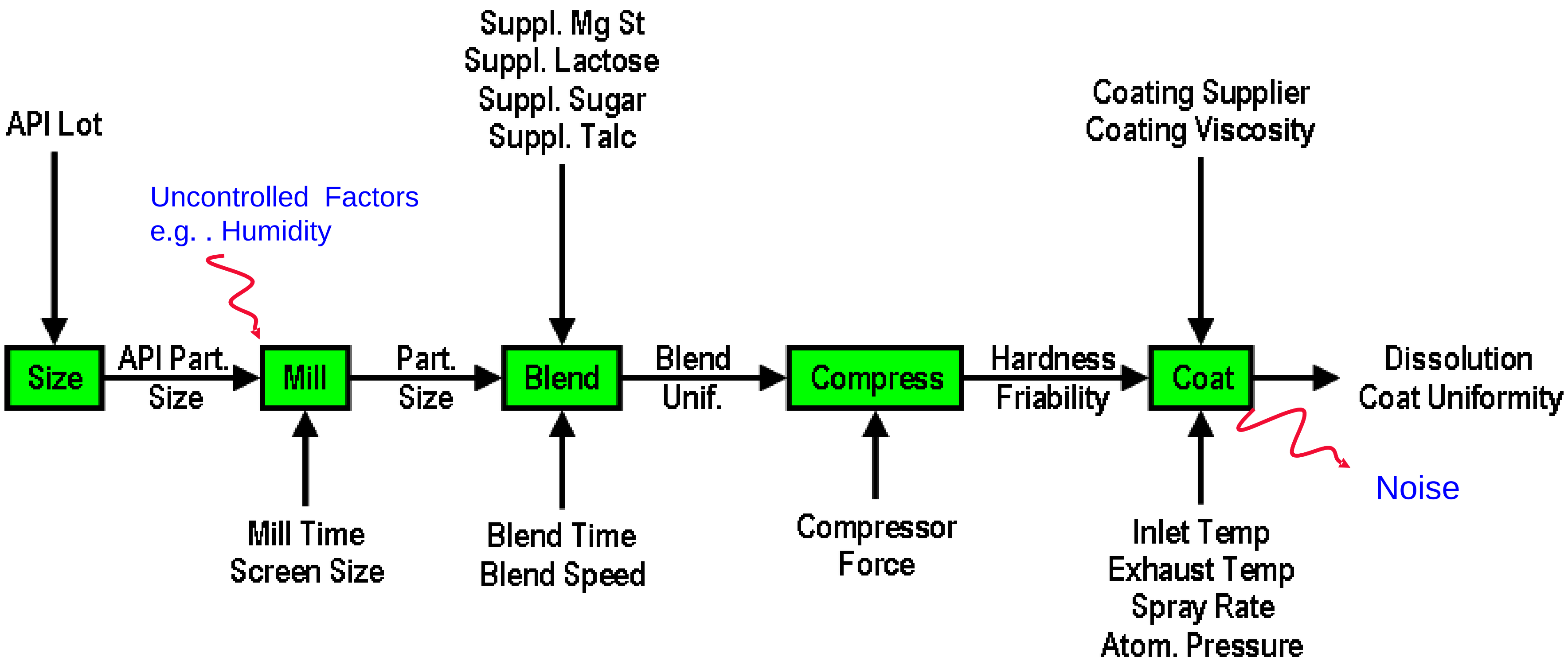


experiment0

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...

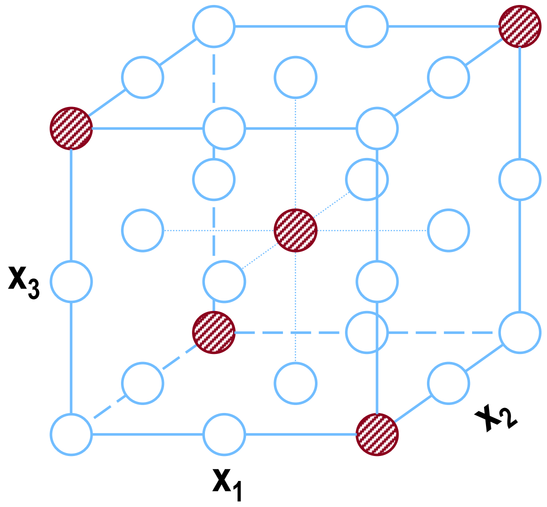
CLASSIC DEFINITION OF DOE

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

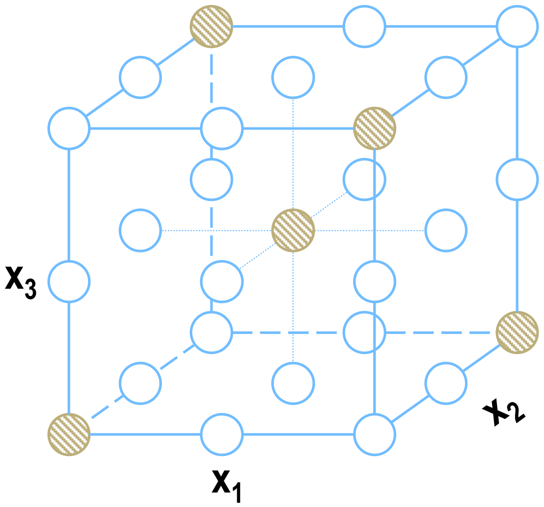


RESPONSE SURFACE DOE IN A NUTSHELL

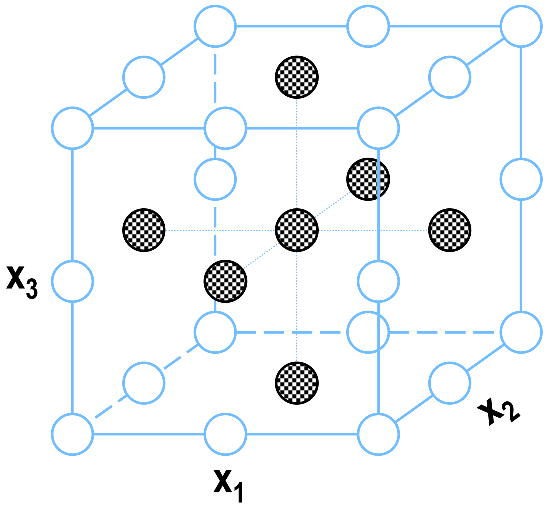
Block 1



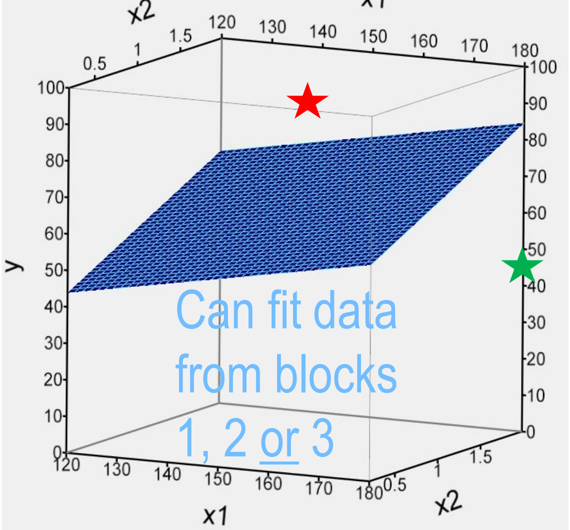
Block 2



Block 3

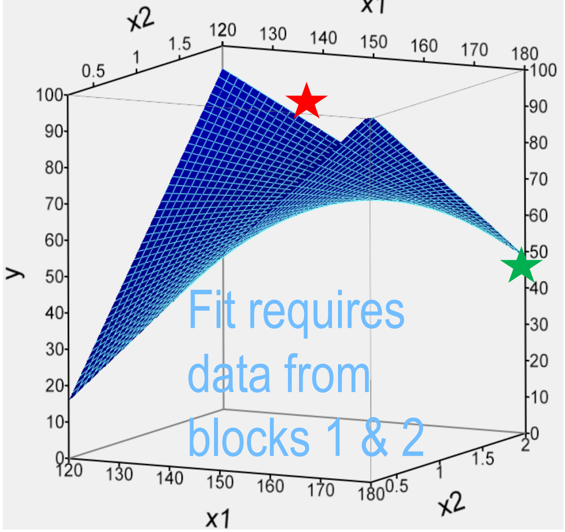


Lack-of-fit

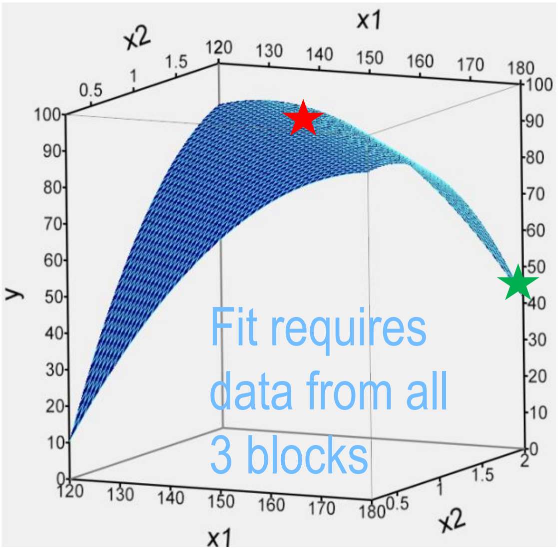


Can fit data from blocks 1, 2 or 3

Lack-of-fit

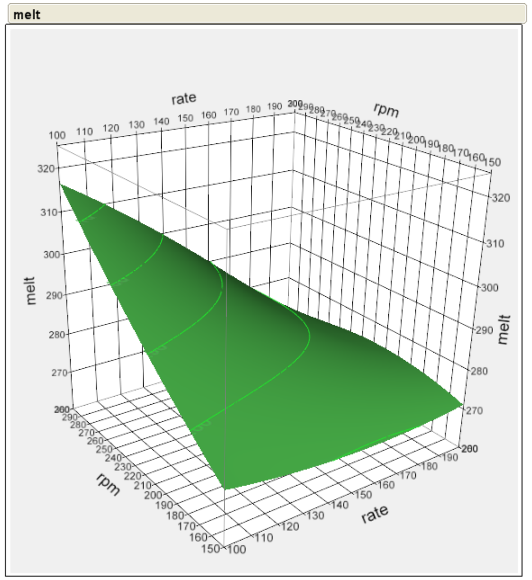


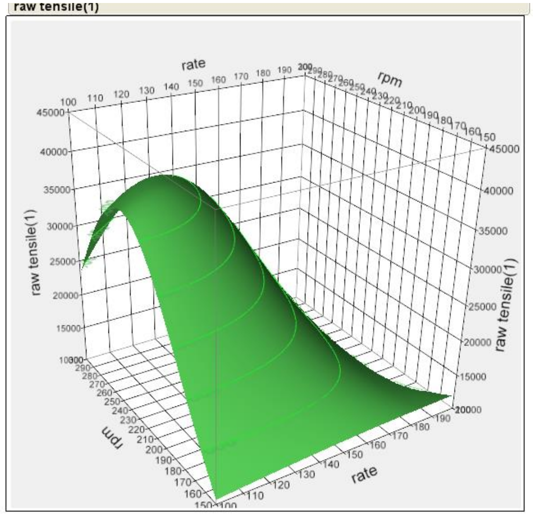
Fit requires data from blocks 1 & 2

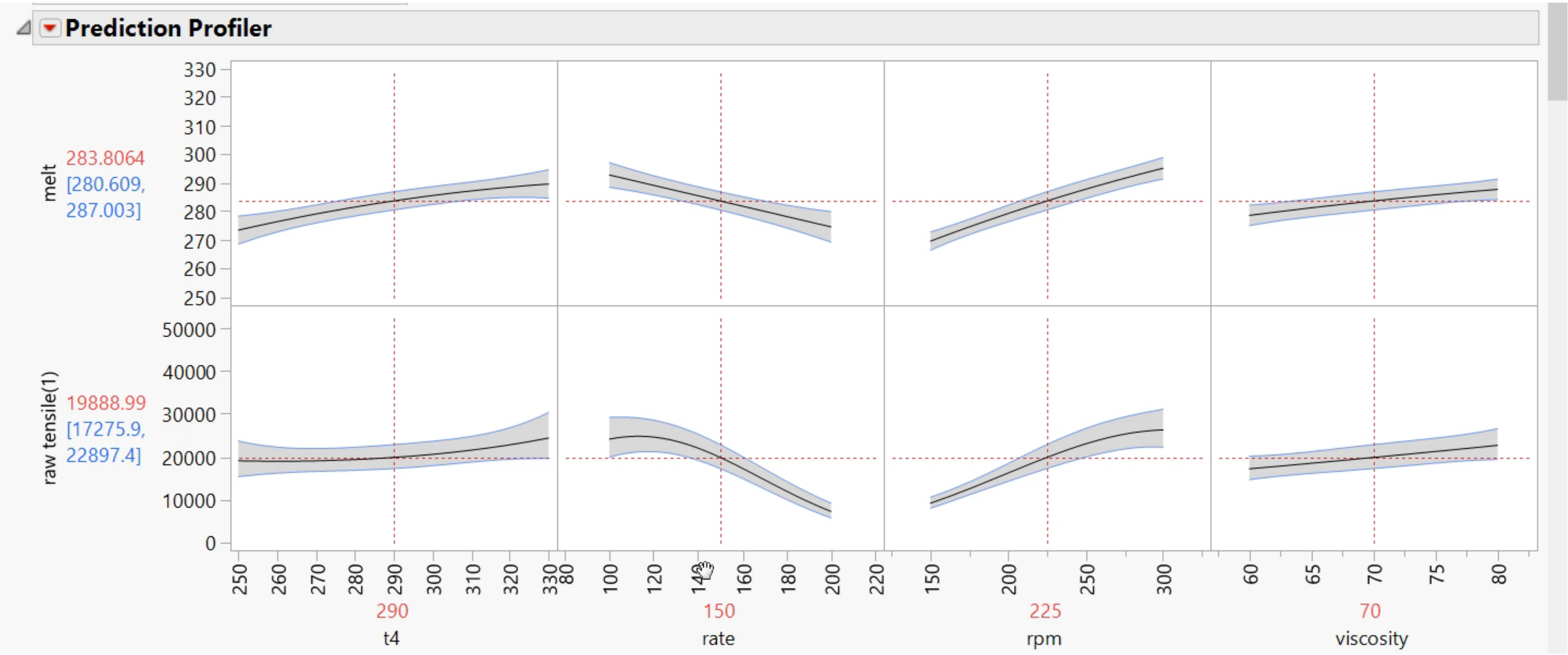


Fit requires data from all 3 blocks

4 CONTROLS (INPUTS) & 2 RESPONSES (OUTPUTS) AND THEIR EMPIRICAL RELATIONSHIPS (MODEL)



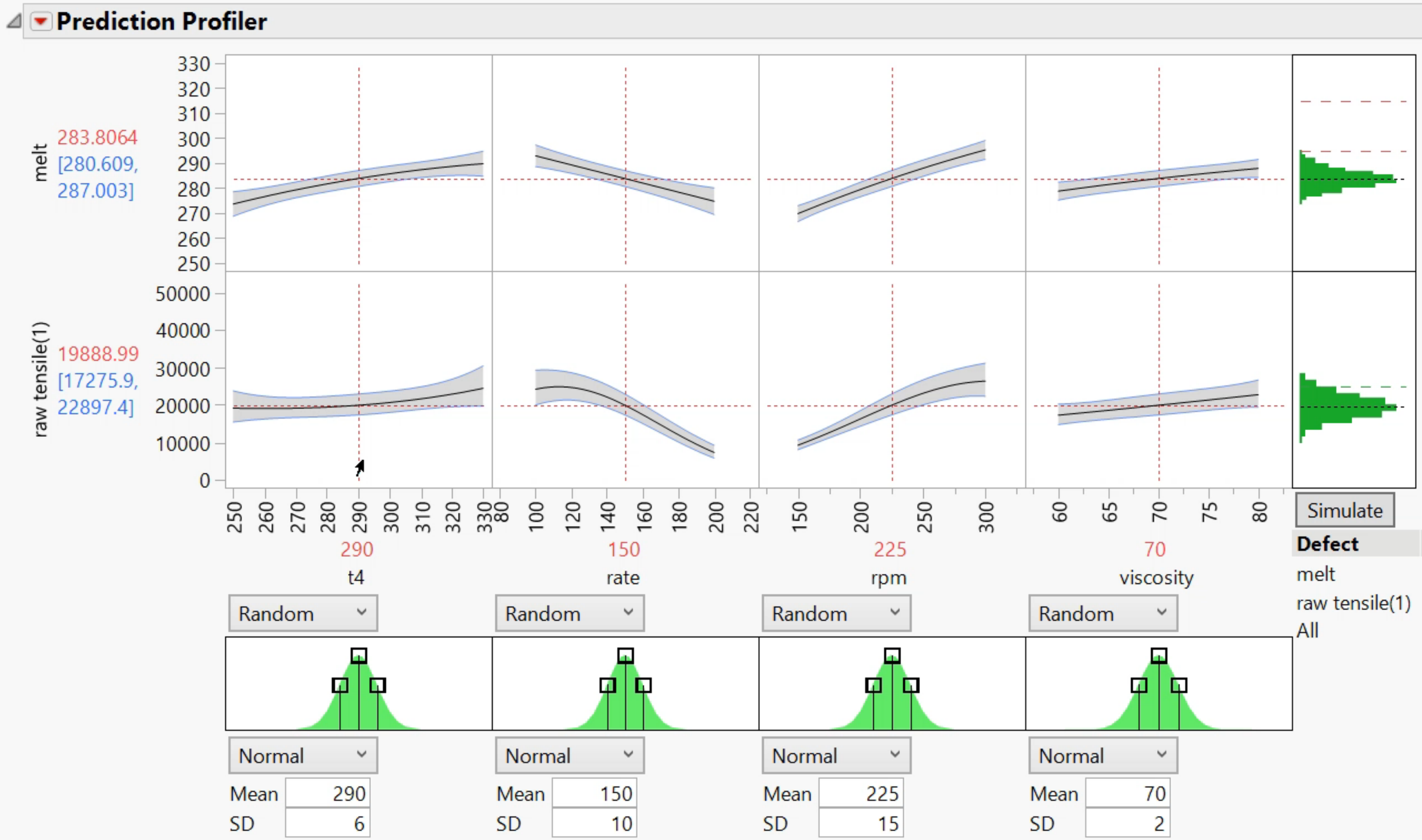




Get these Response Surfaces and Prediction Profiler as result of analyzing data collected for a DOE

ASSESS UNCERTAINTY IN SURROGATE MODEL PREDICTIONS EVEN 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.



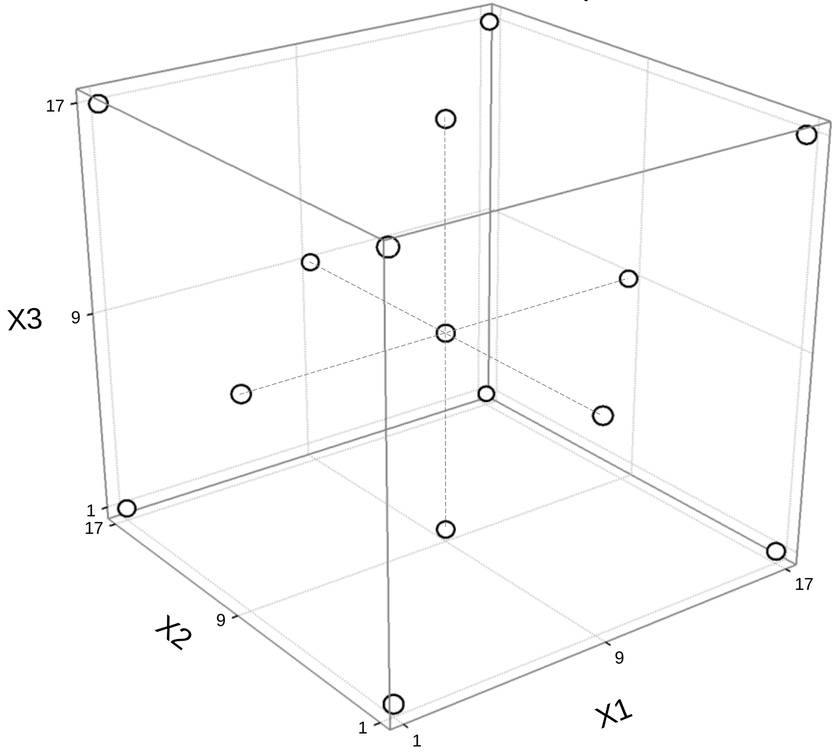
Defect	Rate	Mean	SD
melt	0.998	283.685	3.4853
raw tensile(1)	0.9522	19634.2	3083.21
All	0.998		

TWO CLASSES OF DESIGNS FOR TWO TYPES OF SURROGATE MODELING OF SIMULATIONS

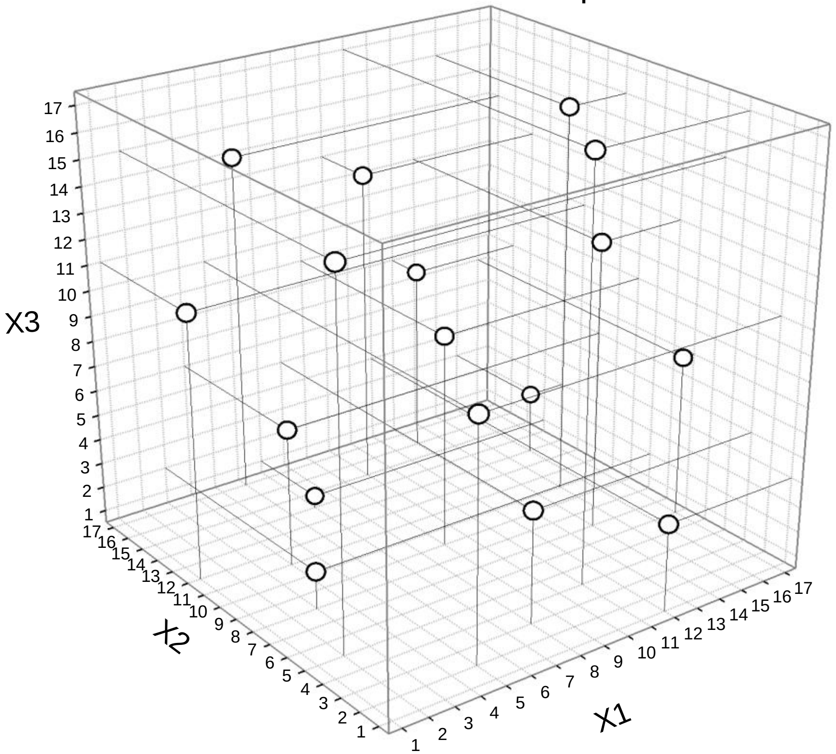
- ***Traditional factorial/response surface*** designs for polynomial modeling with categorical (qualitative) and continuous (quantitative) variables
  - Designs can be sequentially constructed to support increasingly complex models
  - Example featured here reanalyzes a simulation case matrix in which all combinations of 6 variable settings were originally run- a total of 648 = 6 X 3 X 3 X 3 X 2 X 2
  - References on Resolution V, Fractional-Factorial Designs for many (40+) factors
    - Mee, R. W. (2004), **Efficient Two-Level Designs for Estimating Main Effects and Two-Factor Interactions**, *Journal of Quality Technology*, 36, 400-412.
    - Sanchez, S.M. and Sanchez, P.J. (2005), **Very Large Fractional Factorial and Central Composite Designs**, *ACM Transactions on Modeling and Computer Simulation*, Vol. 15, No. 4, October 2005, Pages 362–377.
    - Xu, H. (2009), **Algorithmic Construction of Efficient Fractional Factorial Designs with Large Run Sizes**, *Technometrics*, http://www.stat.ucla.edu/~hqxu/pub/ffd2r3.pdf
- ***Space-filling*** designs primarily for use with continuous and categorical variables AND non-stochastic/deterministic responses
  - These designs can support “Gaussian Process” or “Kriging” spatial regression analysis – an interpolation technique, as well as linear regression – an approximation method

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.

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

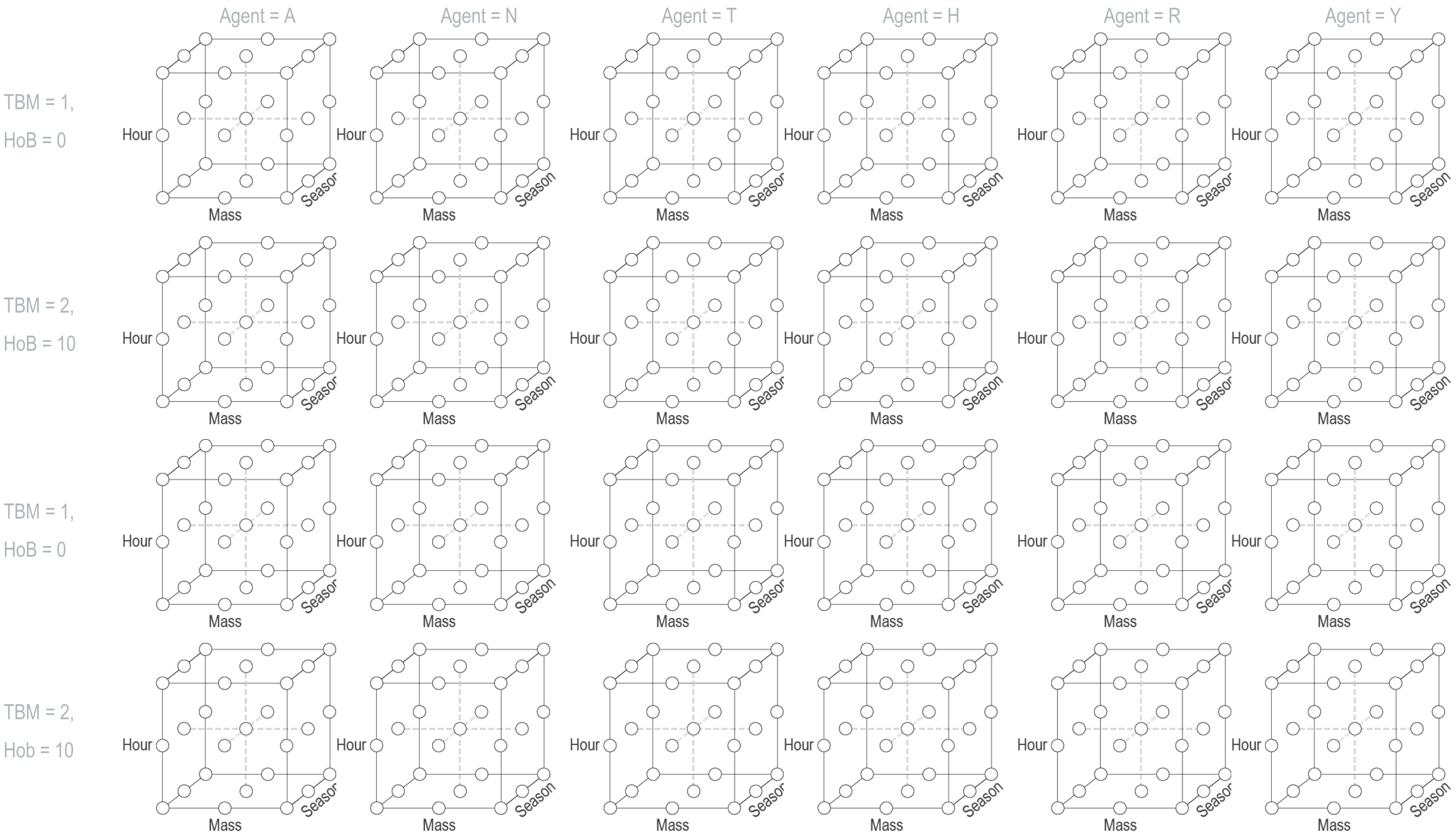
SEQUENTIAL DESIGNS

- 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 (X1)	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 (X2)	2	1 TBM & 1 m, 2 TBMs & 1000 m (categorical)
Mass (relative)	3	1.00, 1.57, 2.00 (continuous)
Height of Burst (X3)	2	0, 10 m (continuous)
Total Cases	648	

ALL 648 POSSIBLE COMBINATIONS OF SETTINGS FOR 6 VARIABLES (6 X 2 X 2 X 3 X 3 X 3)

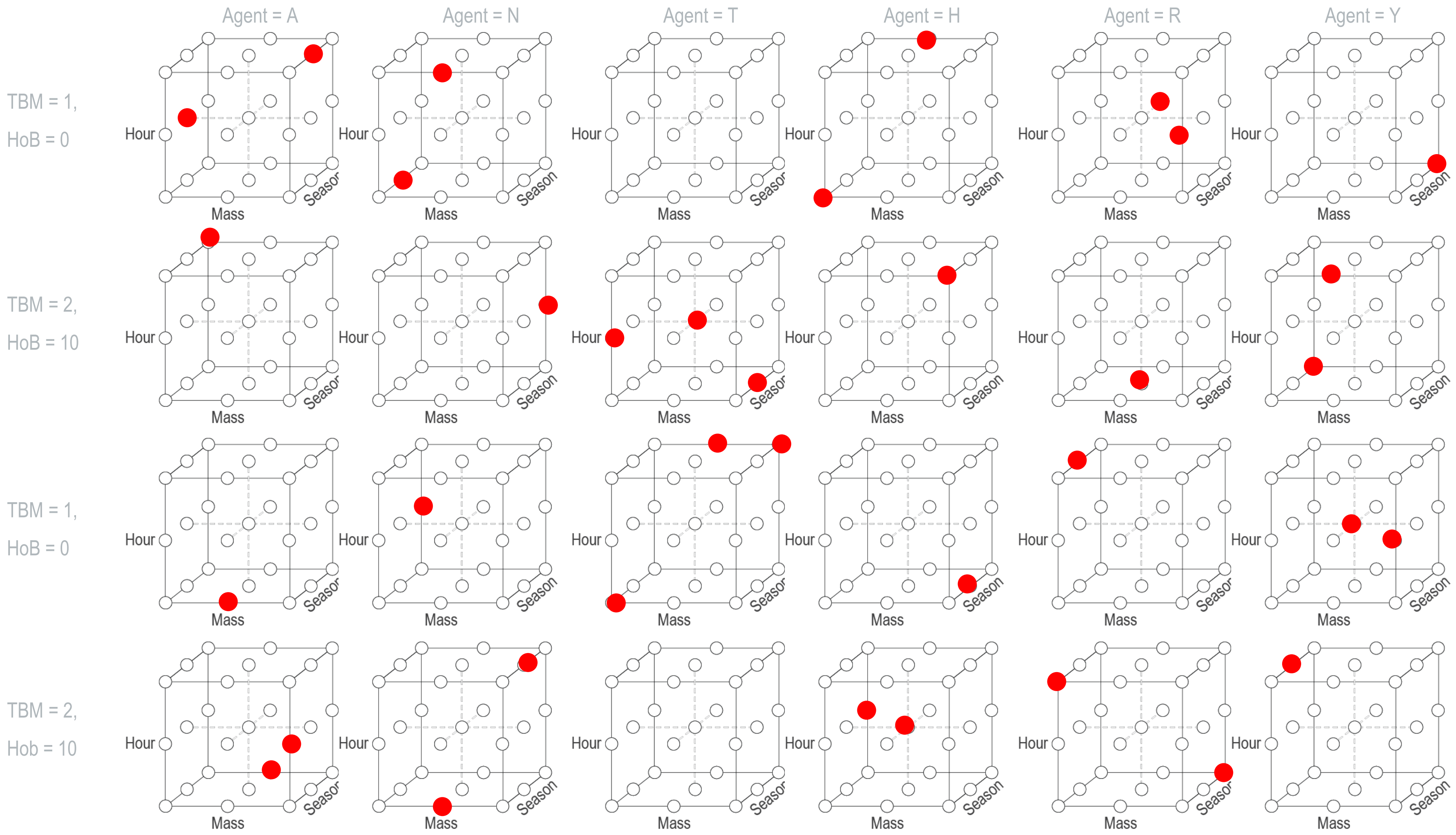


FOUR STAGE DESIGN SEQUENCE

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 →

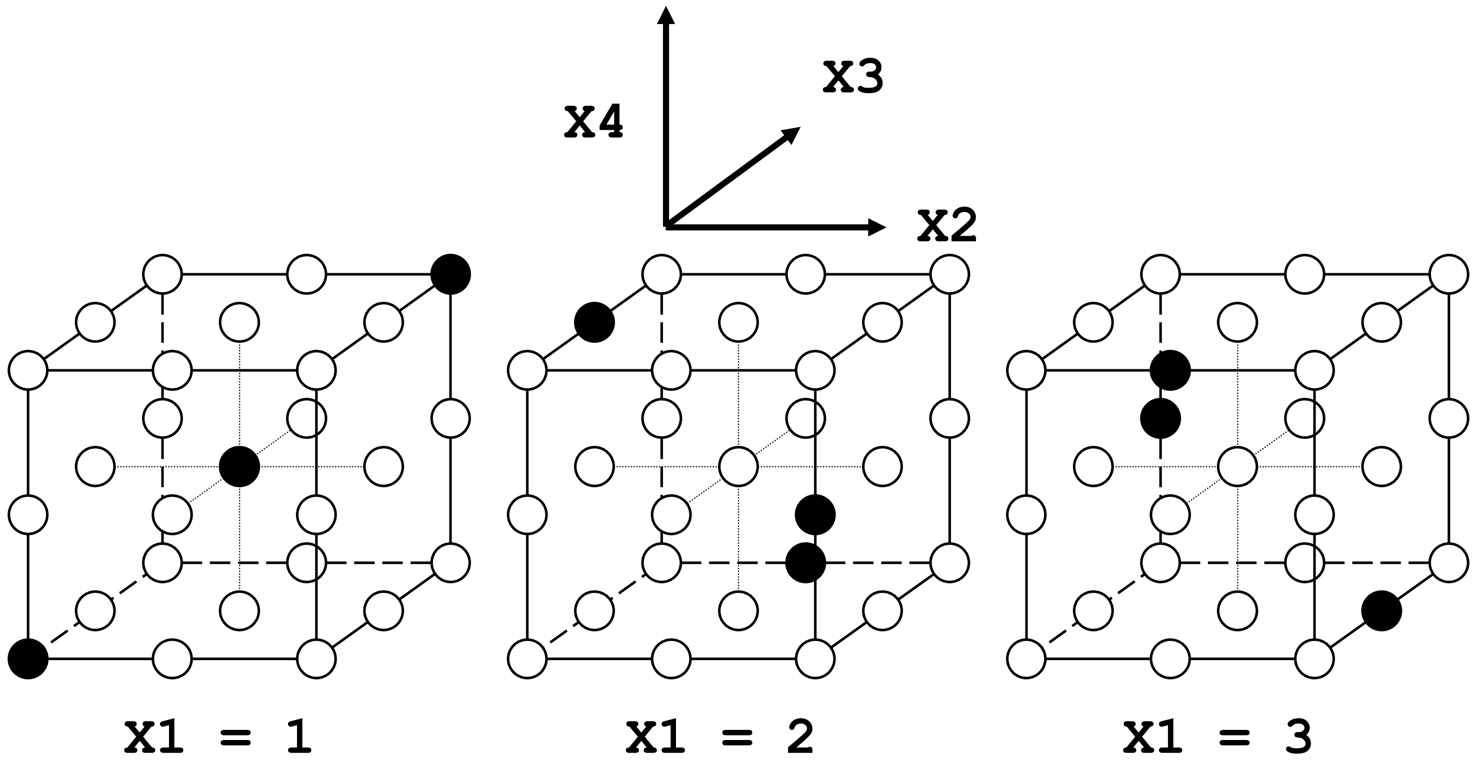
36 OF ALL 648 POSSIBLE COMBINATIONS OF SETTINGS FOR 6 VARIABLES (6 X 2 X 2 X 3 X 3 X 3)



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

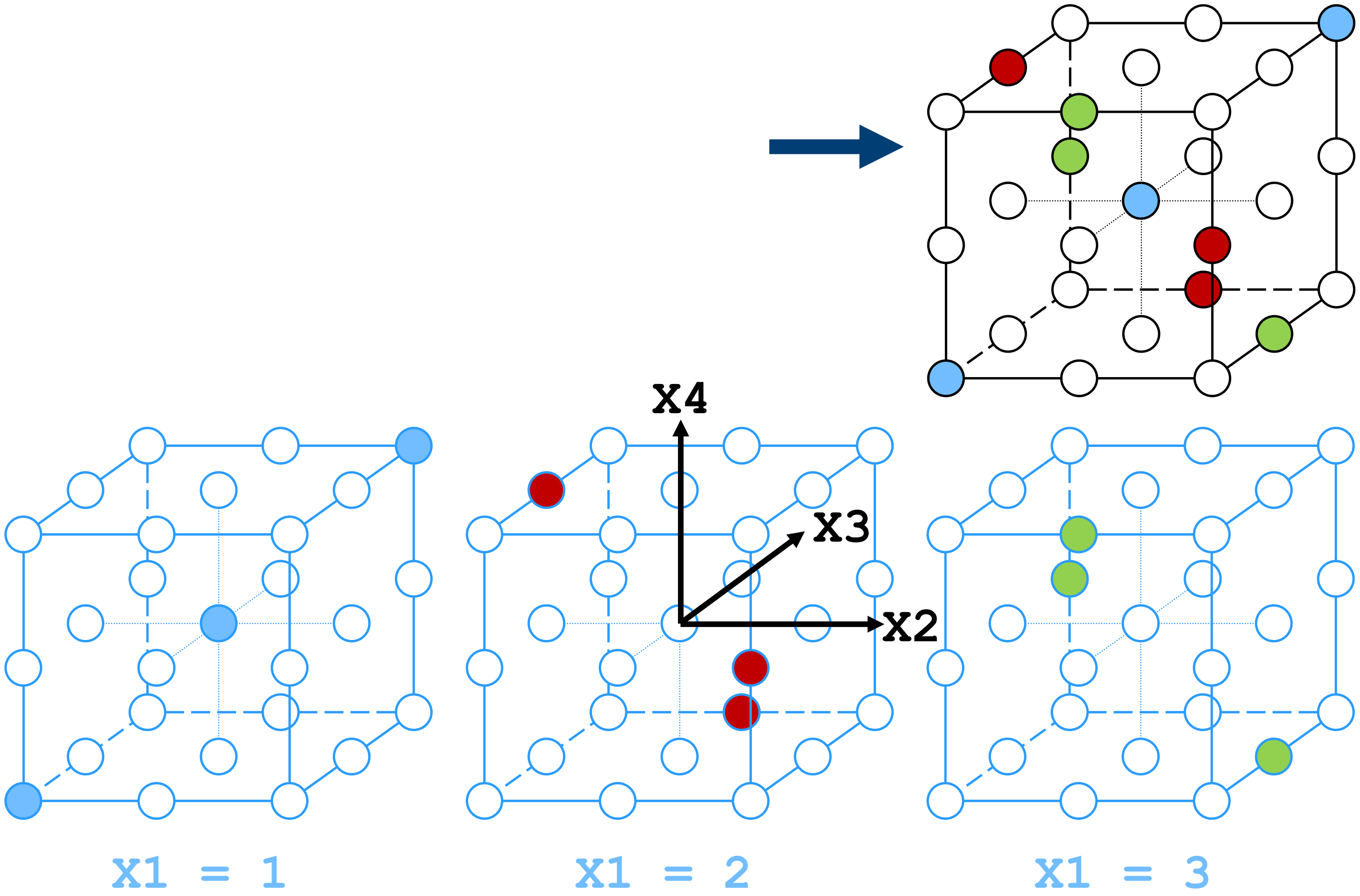
Locations of Trials for a 4-variable, 9-trial Orthogonal Array Design

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

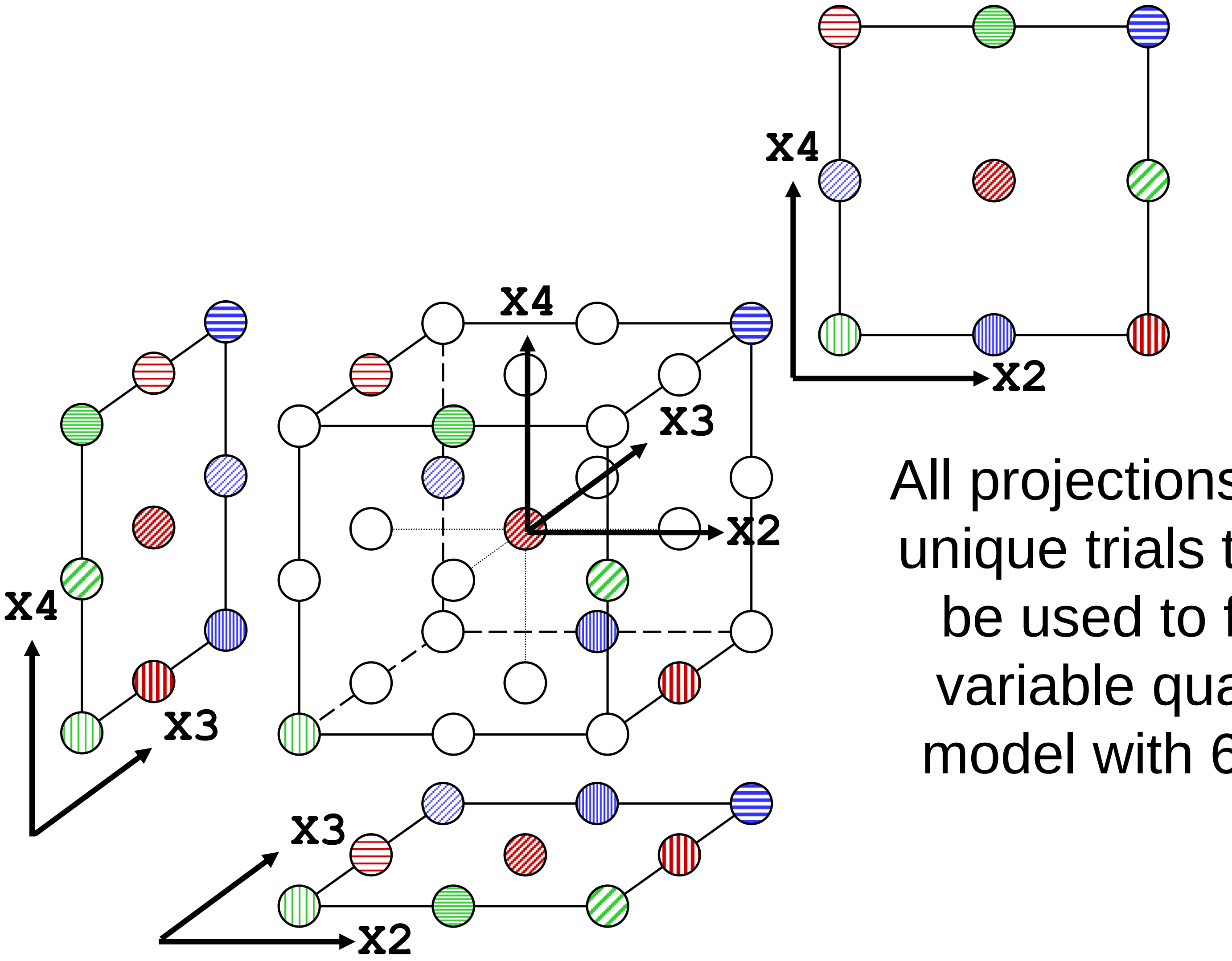


Delete X1 and View Locations of Trials for a 3-Variable OA9 Design

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



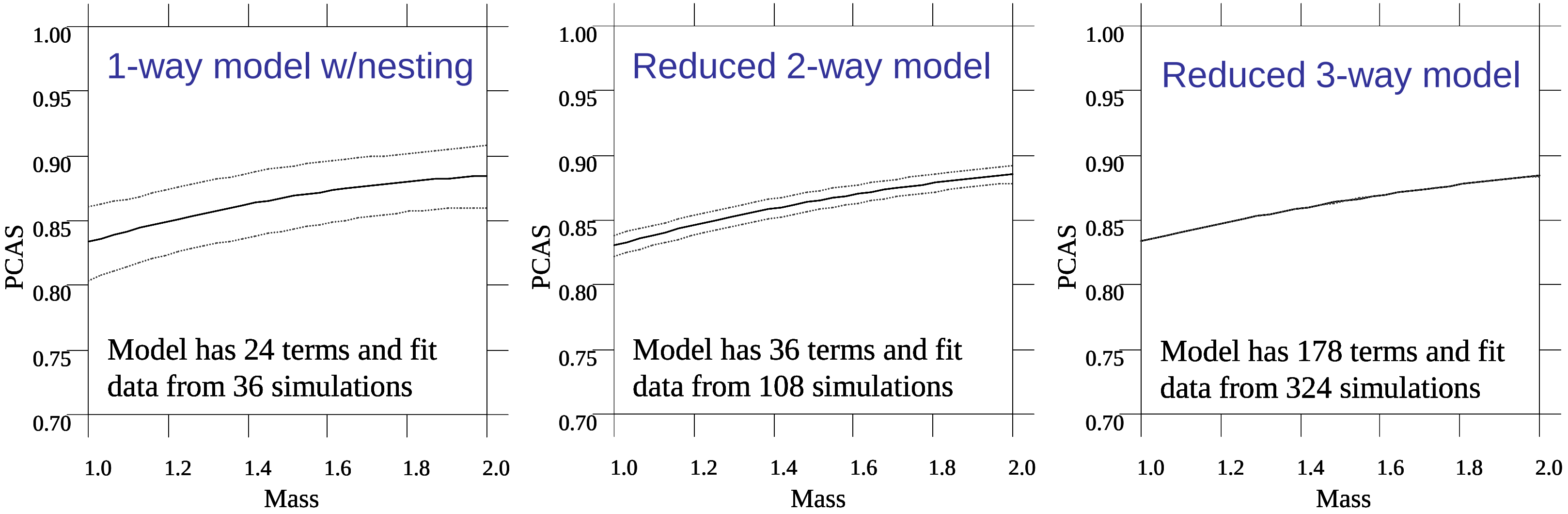
All projections have 9 unique trials that can be used to fit a 2-variable quadratic model with 6 terms

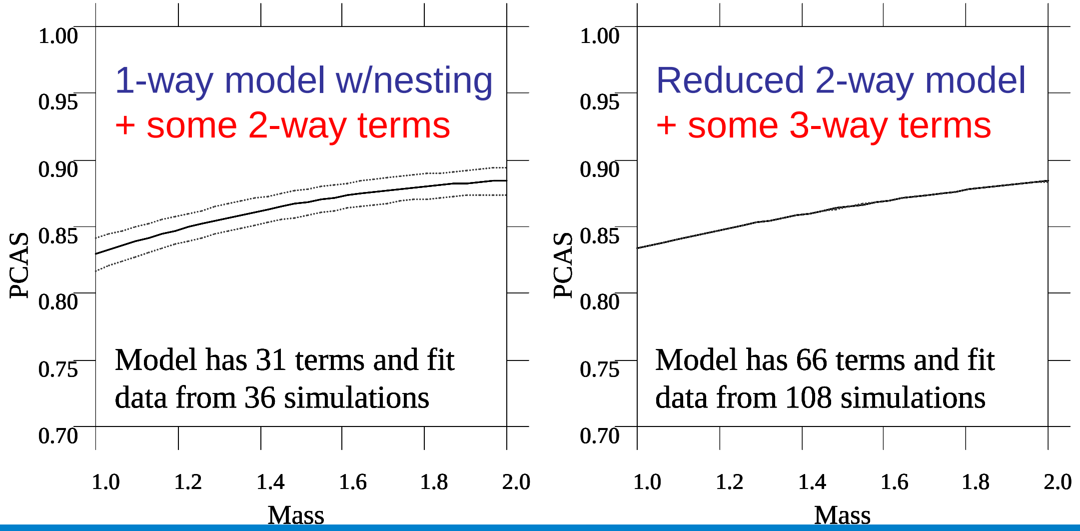
Can Get Designs from Different Sources

- 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

Predicted Probability of Casualty (PCAS) vs. Mass – with Mass Treated as a Continuous Variable – for 5 Different Models Fit to 3 Sets of Simulation Data





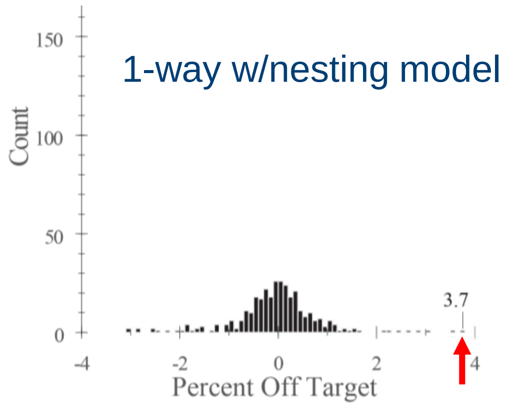
Five other variables were held constant at these settings:

Agent = R
Season = F
Time = 12
HOR = 0
#TBM &
Spread Radius = 1

Predicted Mass
95% Prediction Limits

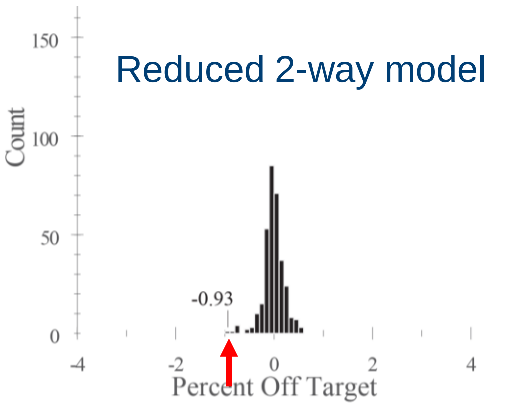
"FACTOR SPARSITY" AND "EFFECT HEREDITY" USED TO ENHANCE MODEL COMPLEXITY

36 trials



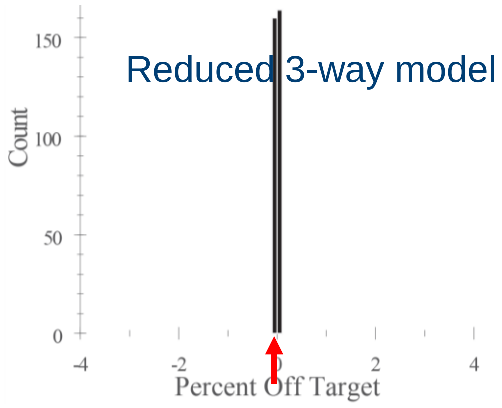
Worst Case = 3.7%
Half of Cases < 0.37%

108 trials

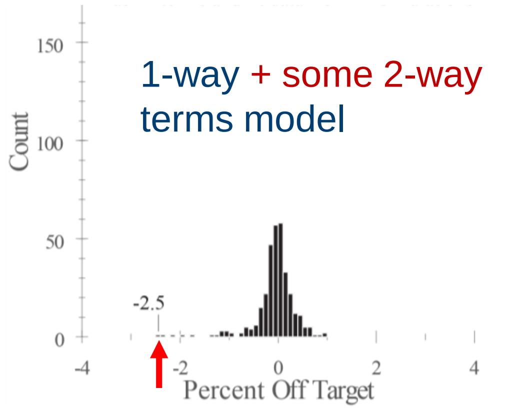


Worst Case = -0.93%
Half of Cases < 0.11%

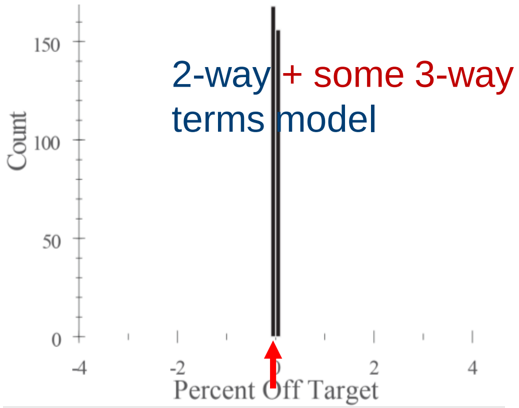
324 trials



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

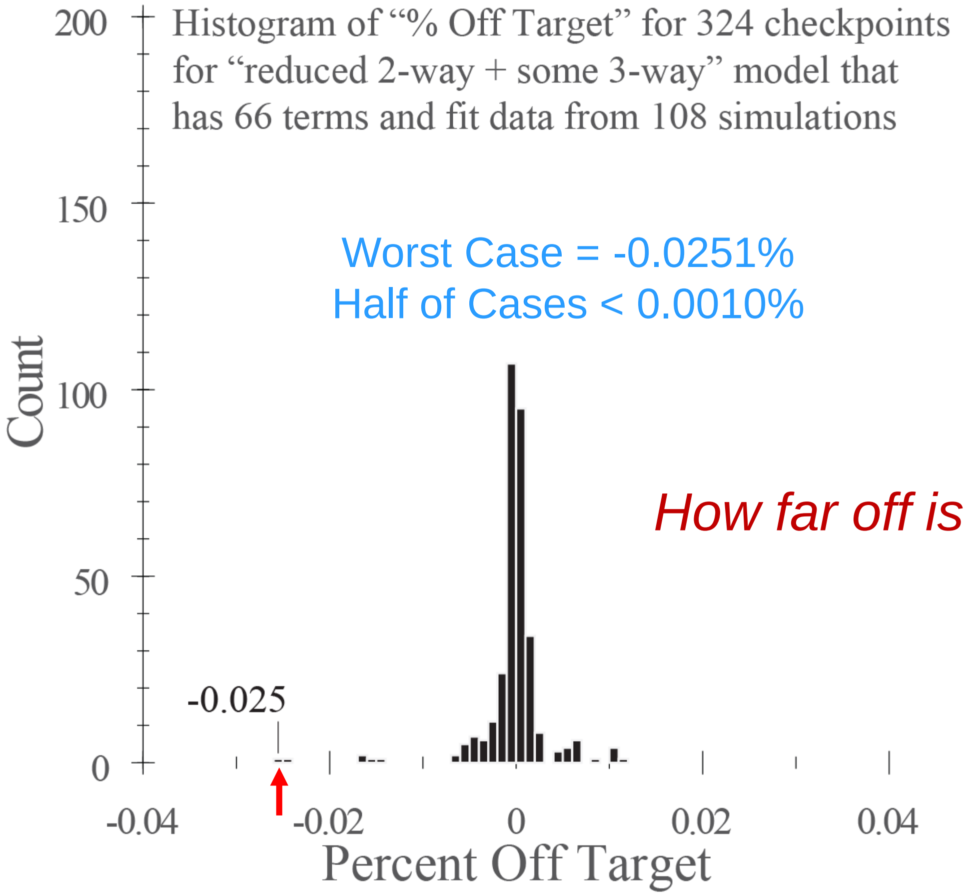
Effect Heredity states significant interactions will only occur if at least one parent is active

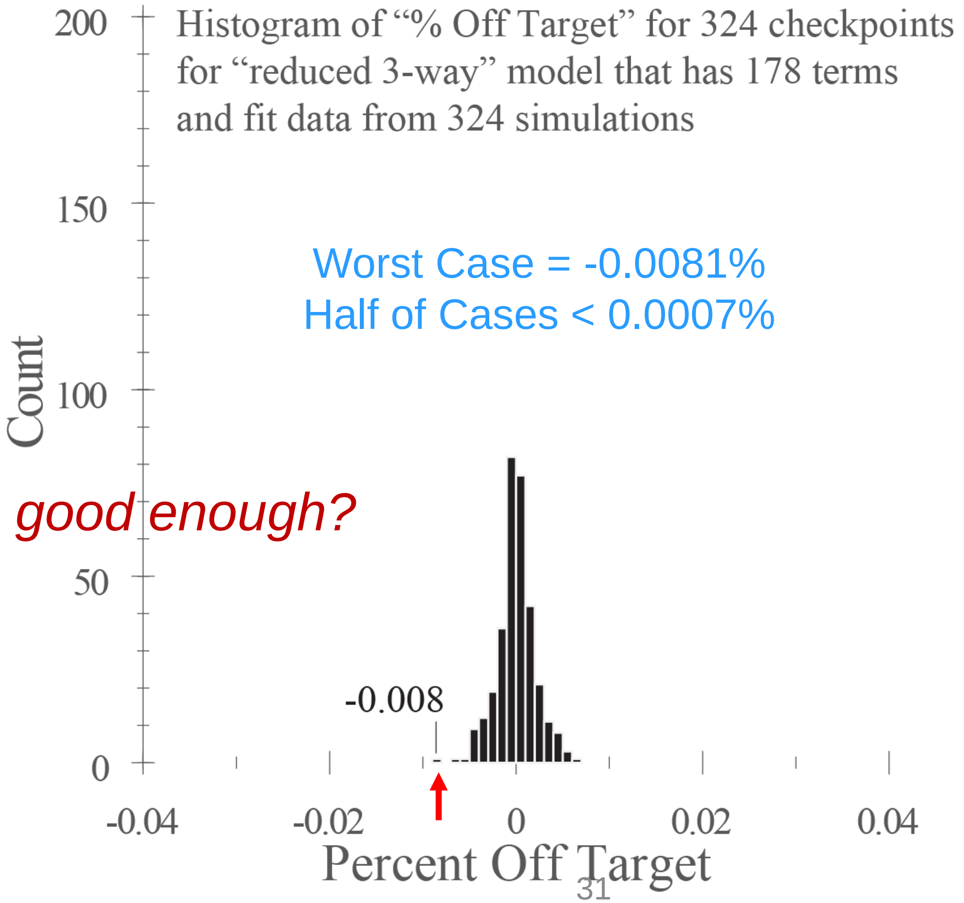
See Wu & Hamada, p. 112

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



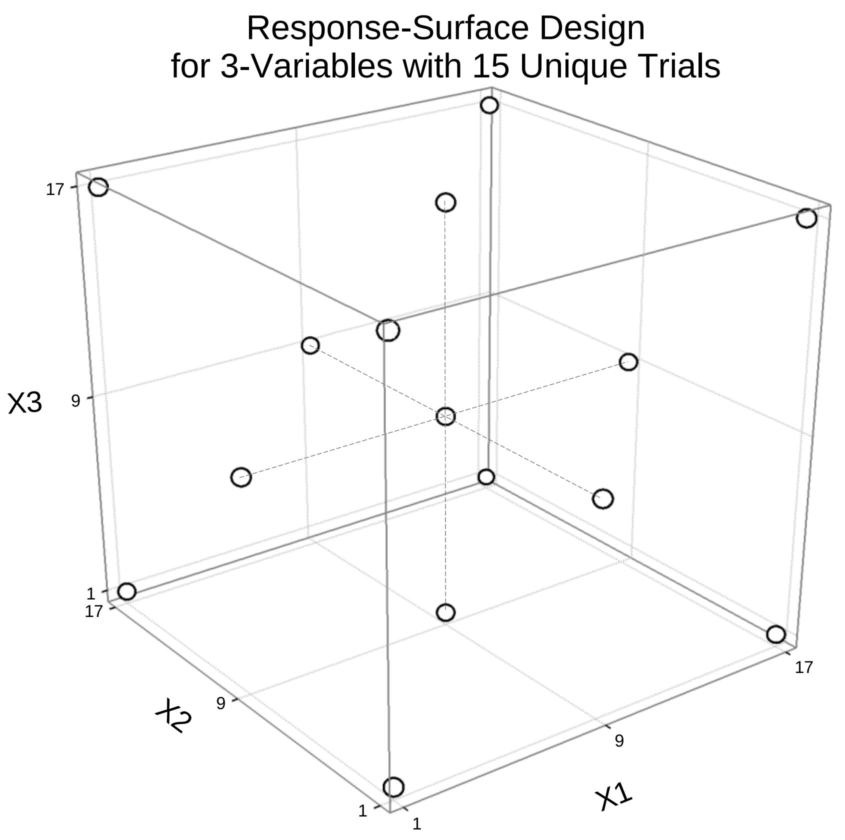


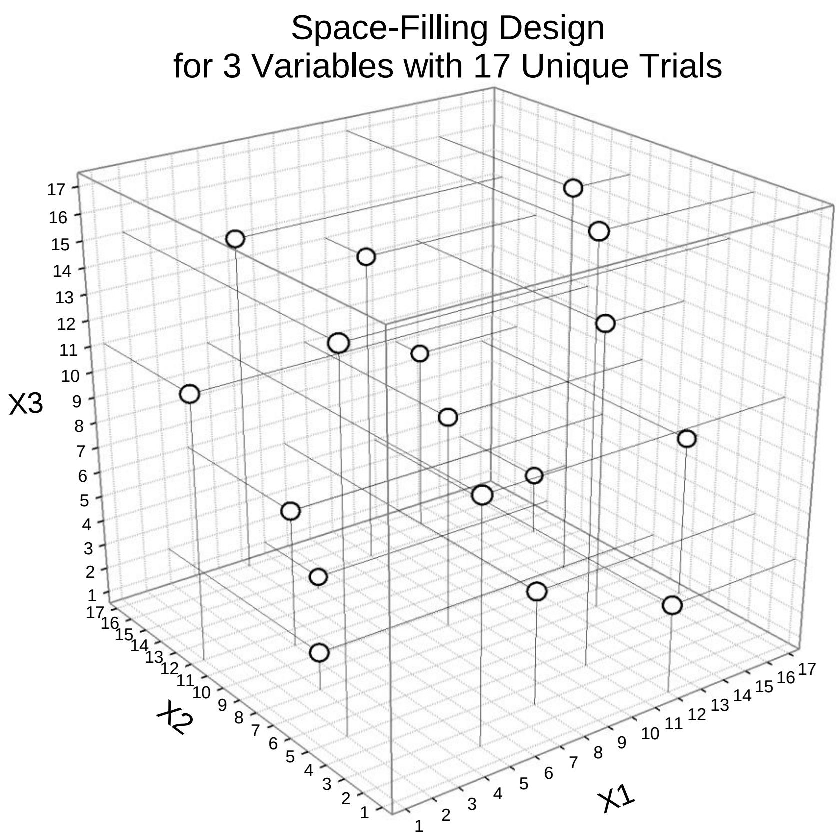
How far off is good enough?

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

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.

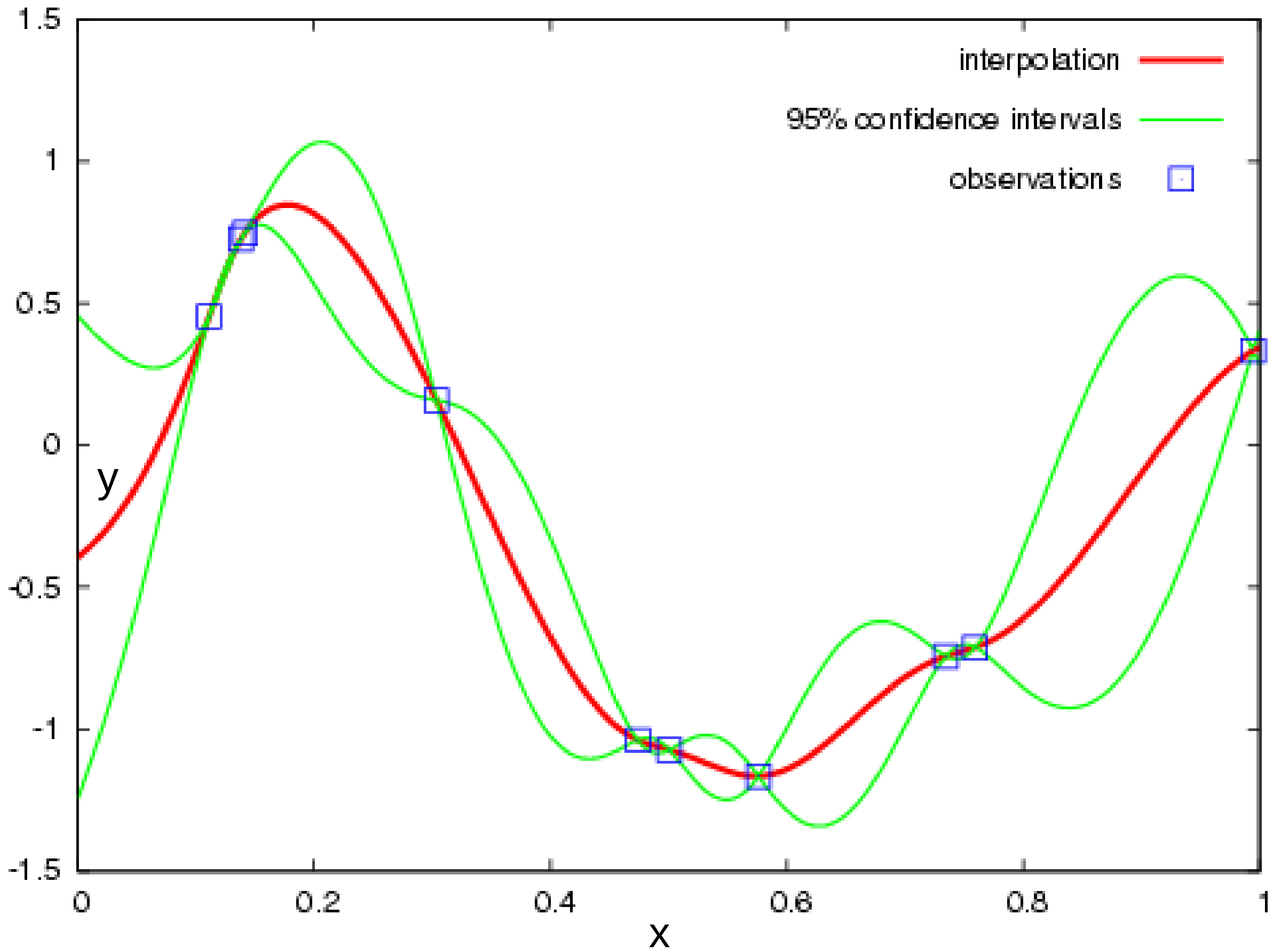
29 CFD SIMULATIONS RUN – 17 USED TO METAMODEL & 12 USED AS CHECKPOINTS

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

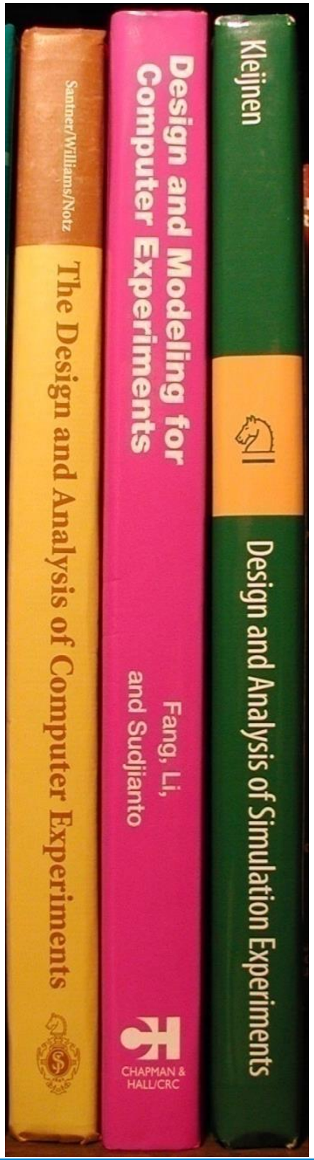
12-trial Plackett-Burman screening design settings used as checkpoints – half just inside and half just outside design boundary (convex hull)

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

KRIGING FIT IN 1-D SHOWING INTERPOLATION AND CONFIDENCE INTERVALS ON PREDICTION



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)

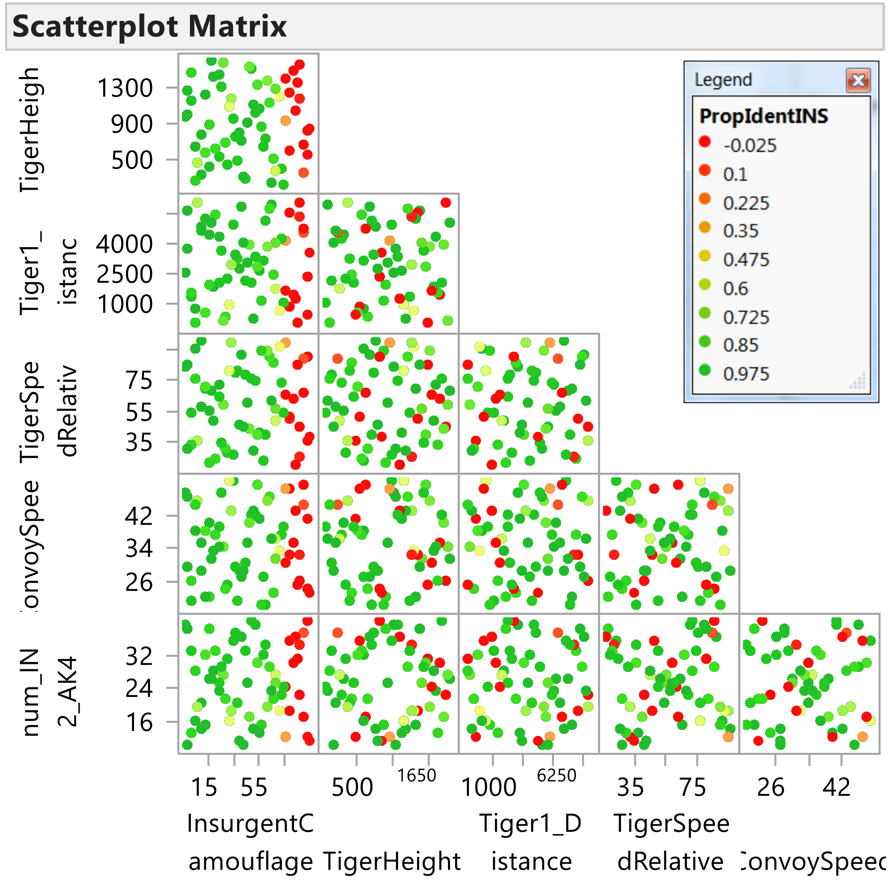
WEBSITES FOR DESIGNS, SOFTWARE & PUBLICATIONS

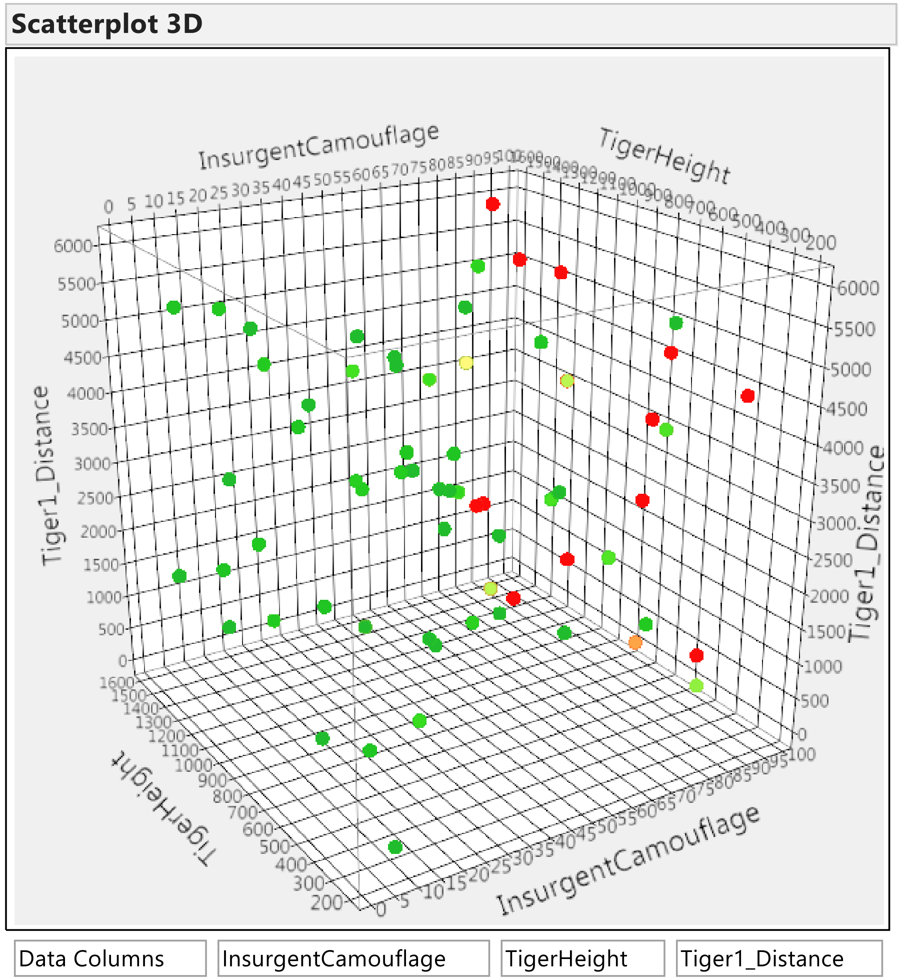
- http://harvest.nps.edu/ The Simulation Experiments & Efficient Design (SEED) Center for Data Farming at Naval Postgraduate School
  - Designs
    - Nearly Orthogonal Latin Hypercubes (NOLH) and
    - Resolution V, Fractional Factorials for many factors
  - Agent-Based Simulation Software
    - Pythagoras
    - 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

SURROGATE MODELING OF A COMPUTER SIMULATION HELICOPTER SURVEILLANCE – IDENTIFYING INSURGENTS

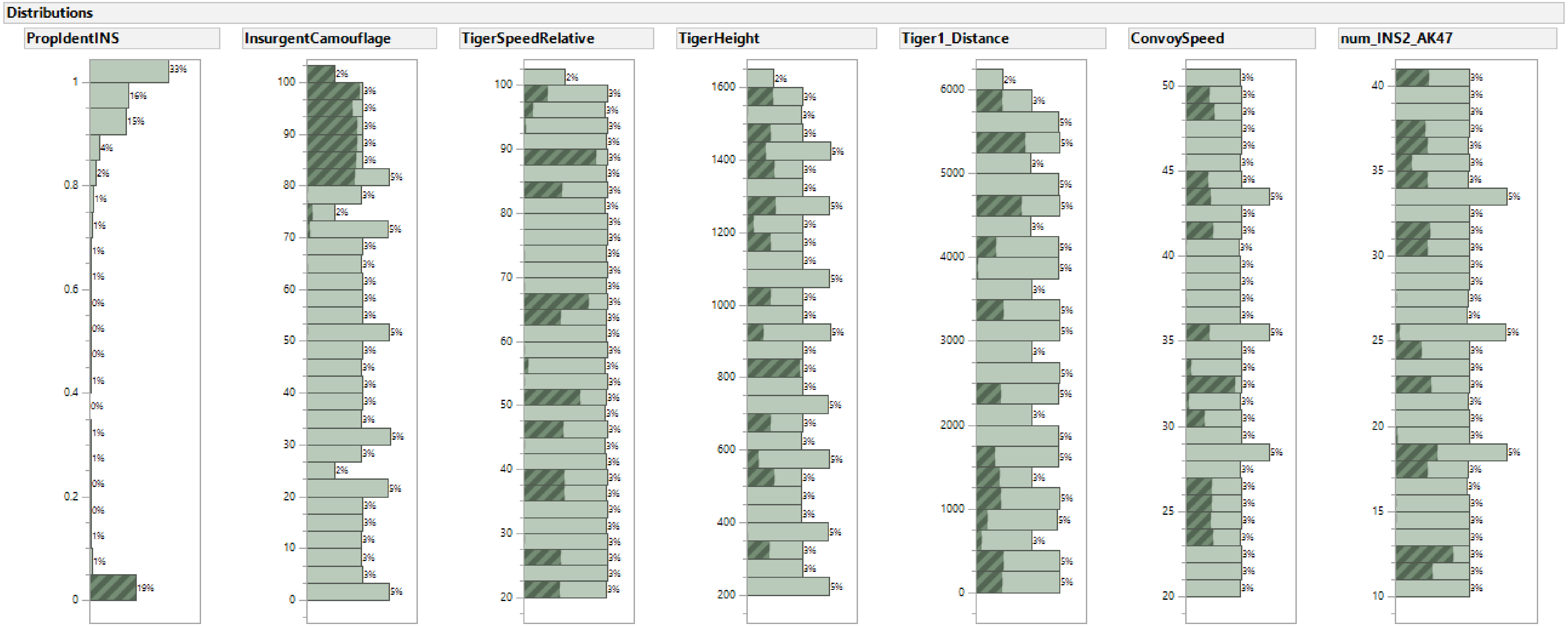
- 2009 International Data Farming Workshop - IDFW21, Lisbon, Portugal
- Largely German team (6 of 8) – their simulation
- 6500 simulations run overnight on cluster in Frankfurt
  - Space Filling Design of Experiments (DOE)
  - 65 unique combinations of 6 factors (each factor at 65 levels)
  - each case had 96 to 100 replications (lost a few)
- Response = Proportion of Insurgents Identified = *PropIdentINS* Data bounded between 0 and 1
- 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

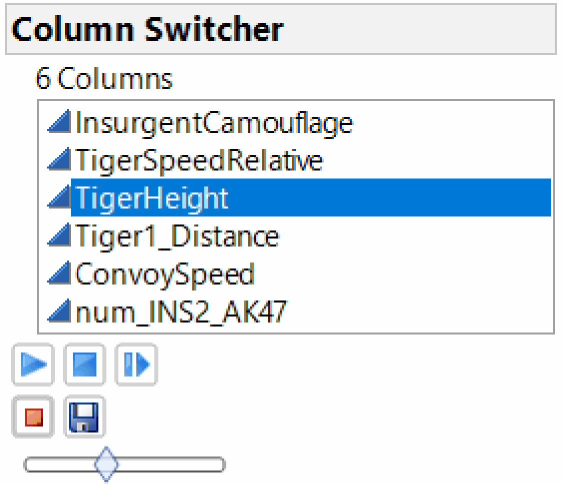
SPACE-FILLING DOE (LATIN HYPERCUBE) VISUALIZED WITH 2-D SCATTERPLOT MATRIX AND 3-D SCATTERPLOT



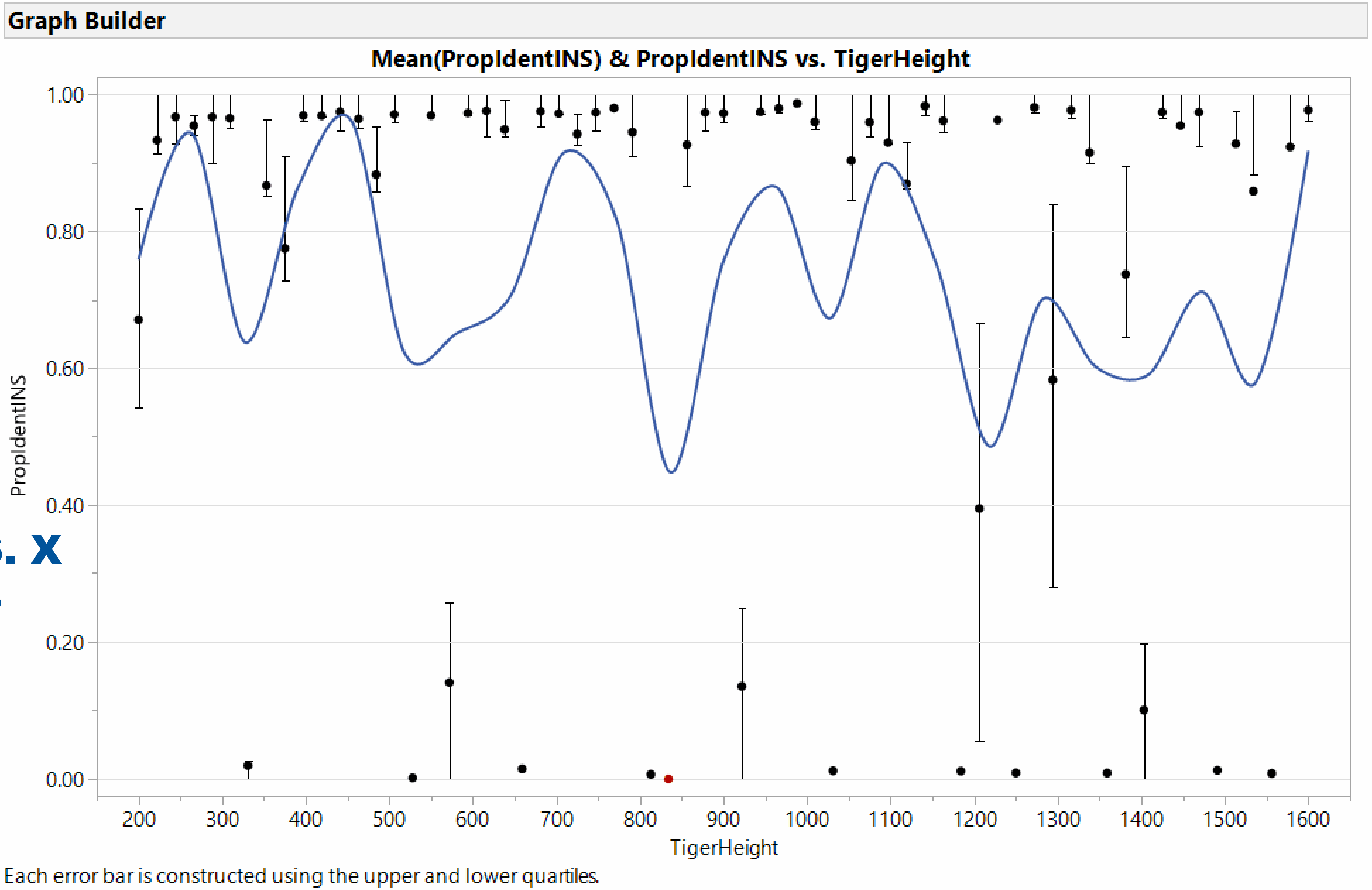


DISTRIBUTIONS OF 1 RESPONSE AND 6 FACTORS

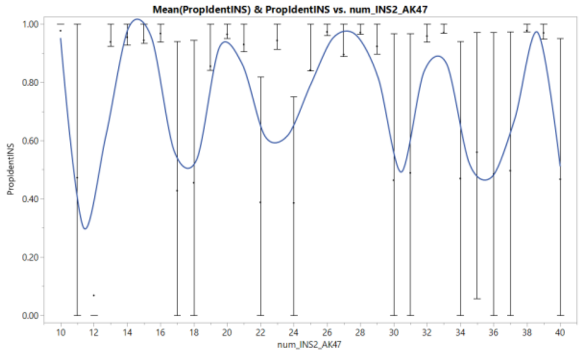


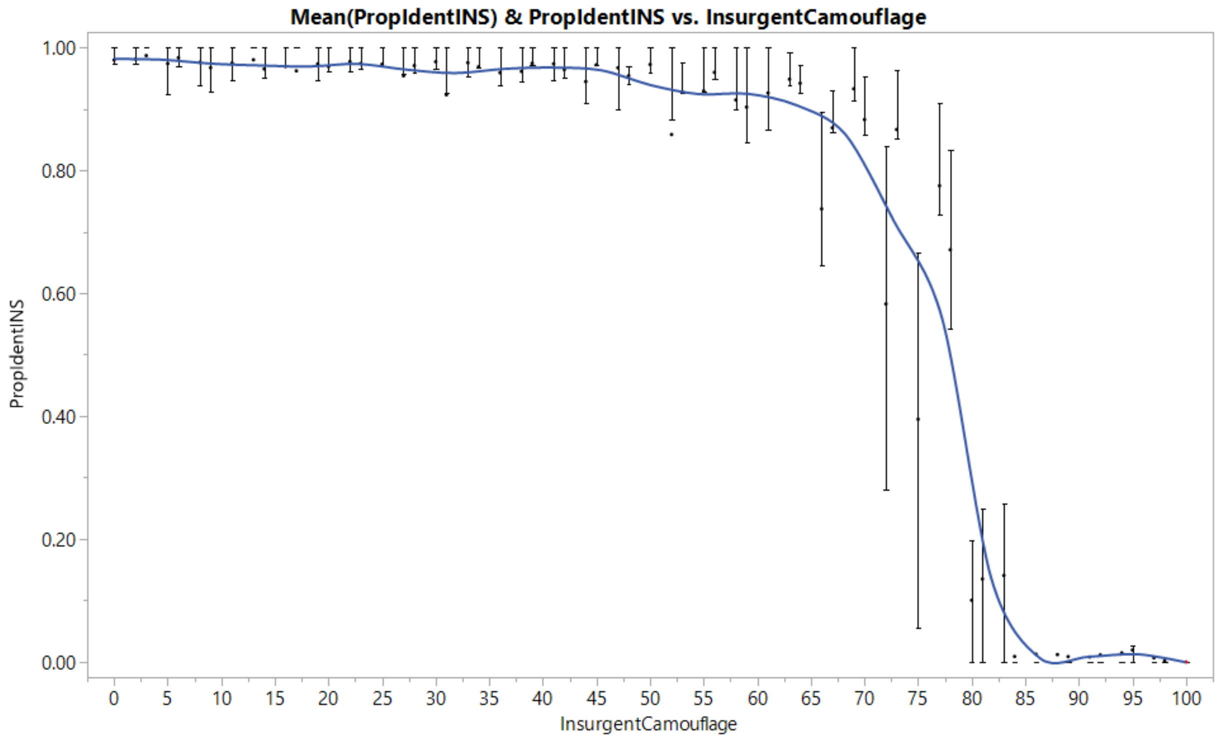


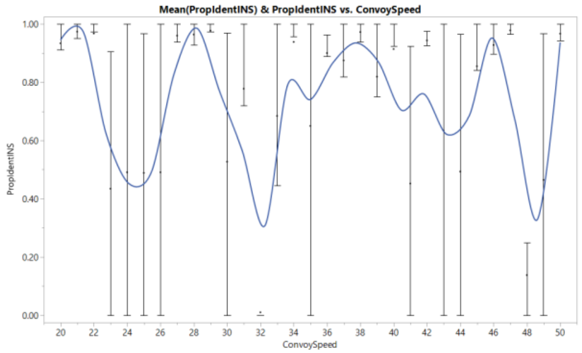
PROPIDENTINS VS. X FOR 6 FACTORS

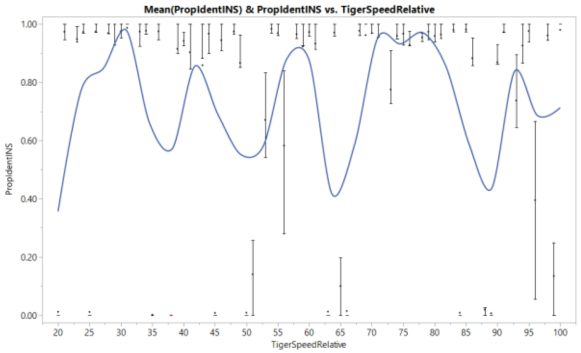


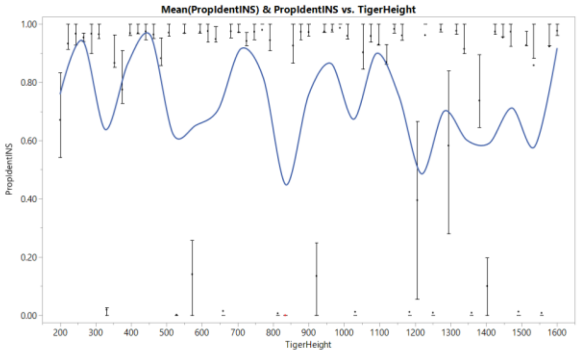
PROPIDENTINS VS. X FOR 6 FACTORS

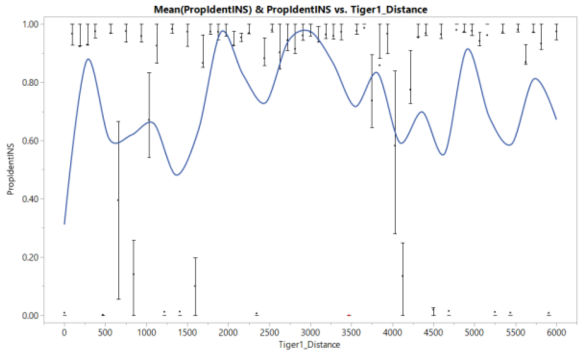




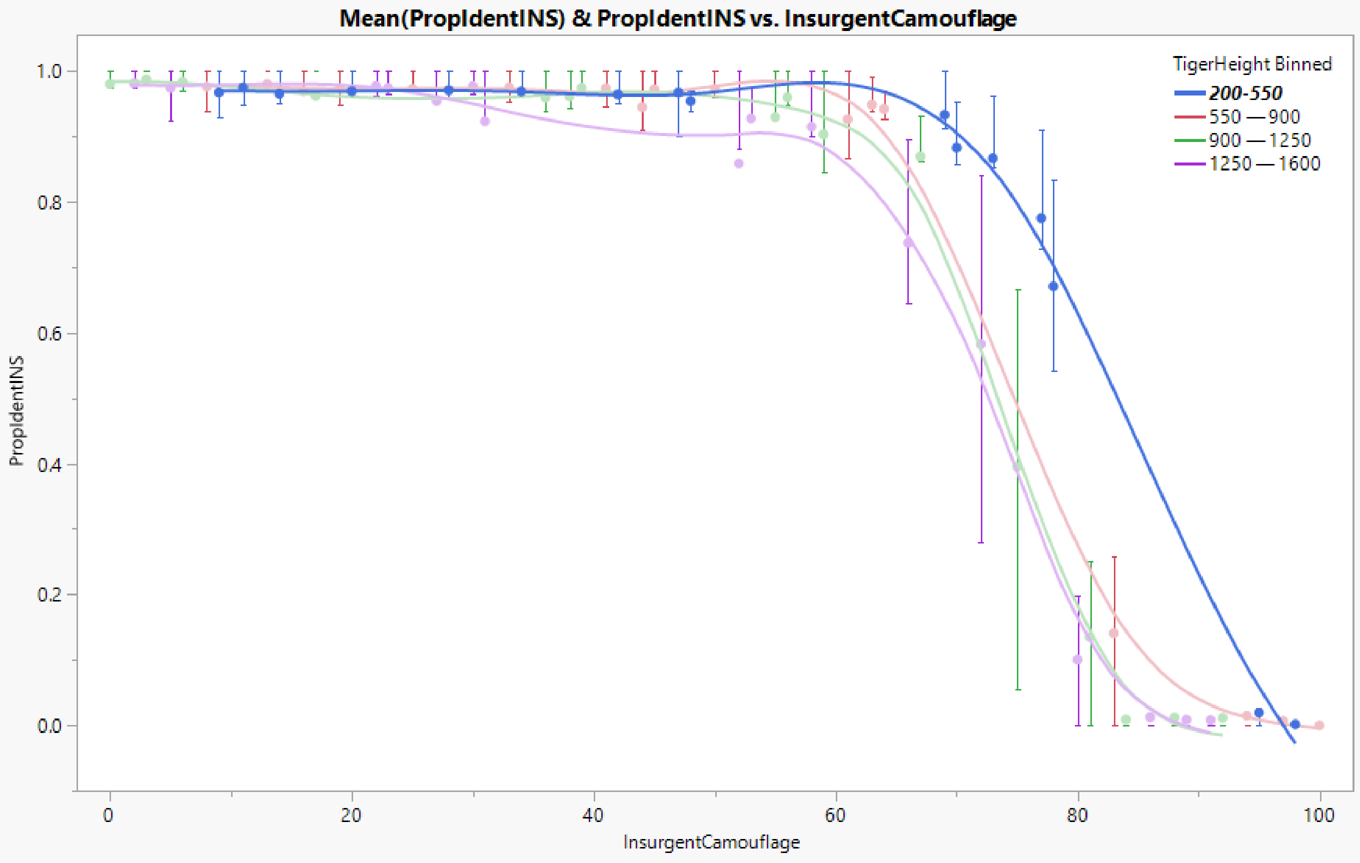


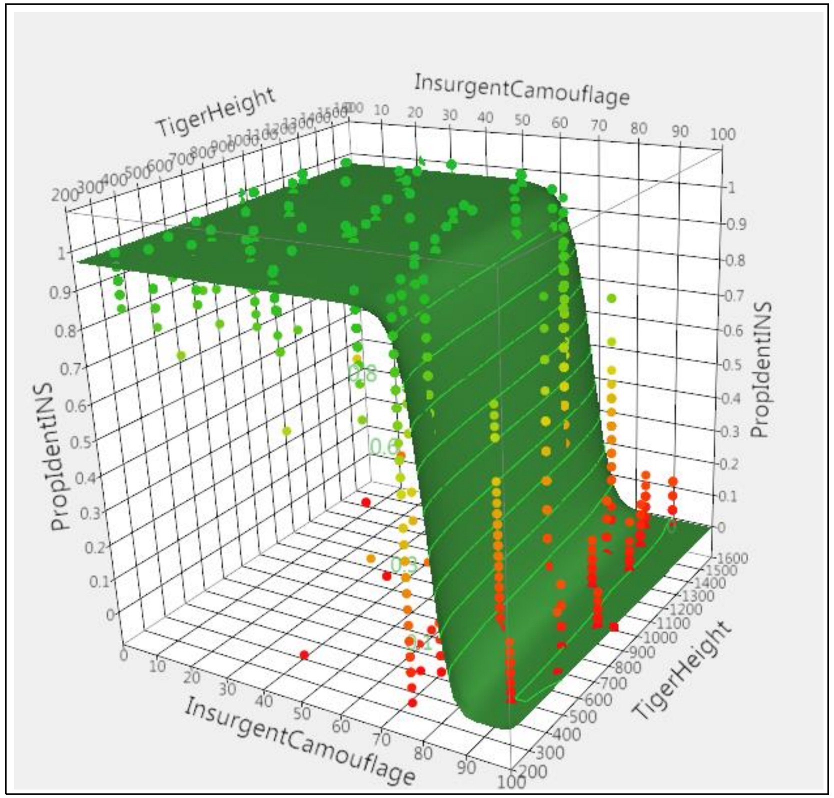






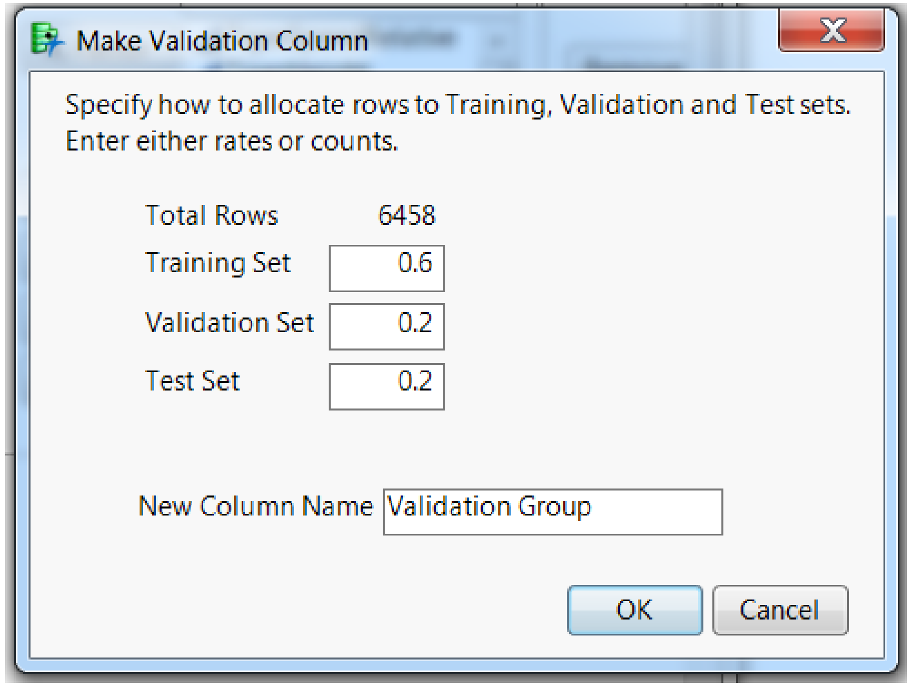
PROPIDENTINS VS. CAMOUFLAGE AT DIFFERENT HEIGHTS

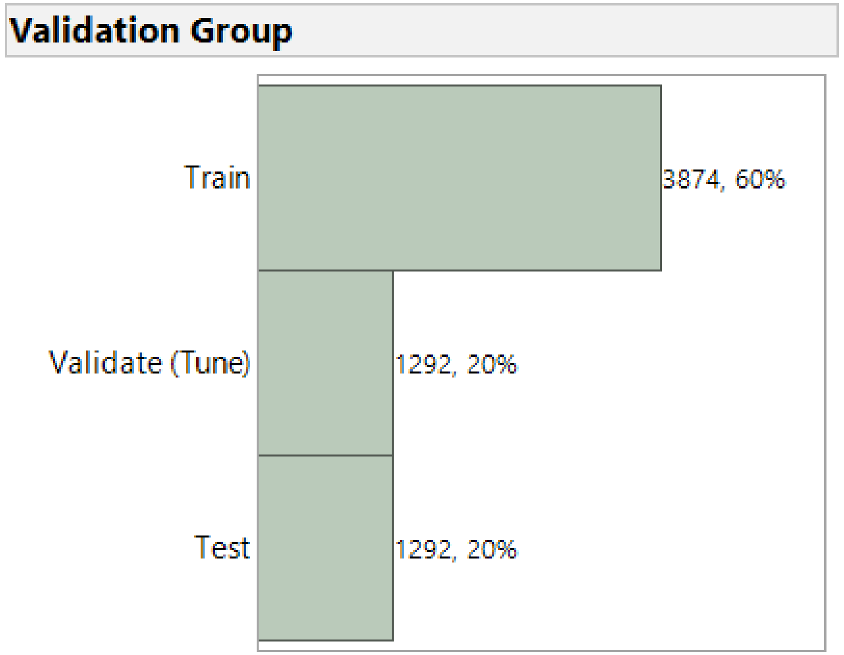




HONEST ASSESSMENT APPROACH
USING TRAIN, VALIDATE (TUNE), AND TEST SUBSETS

Used in model selection and estimating its prediction error on new data



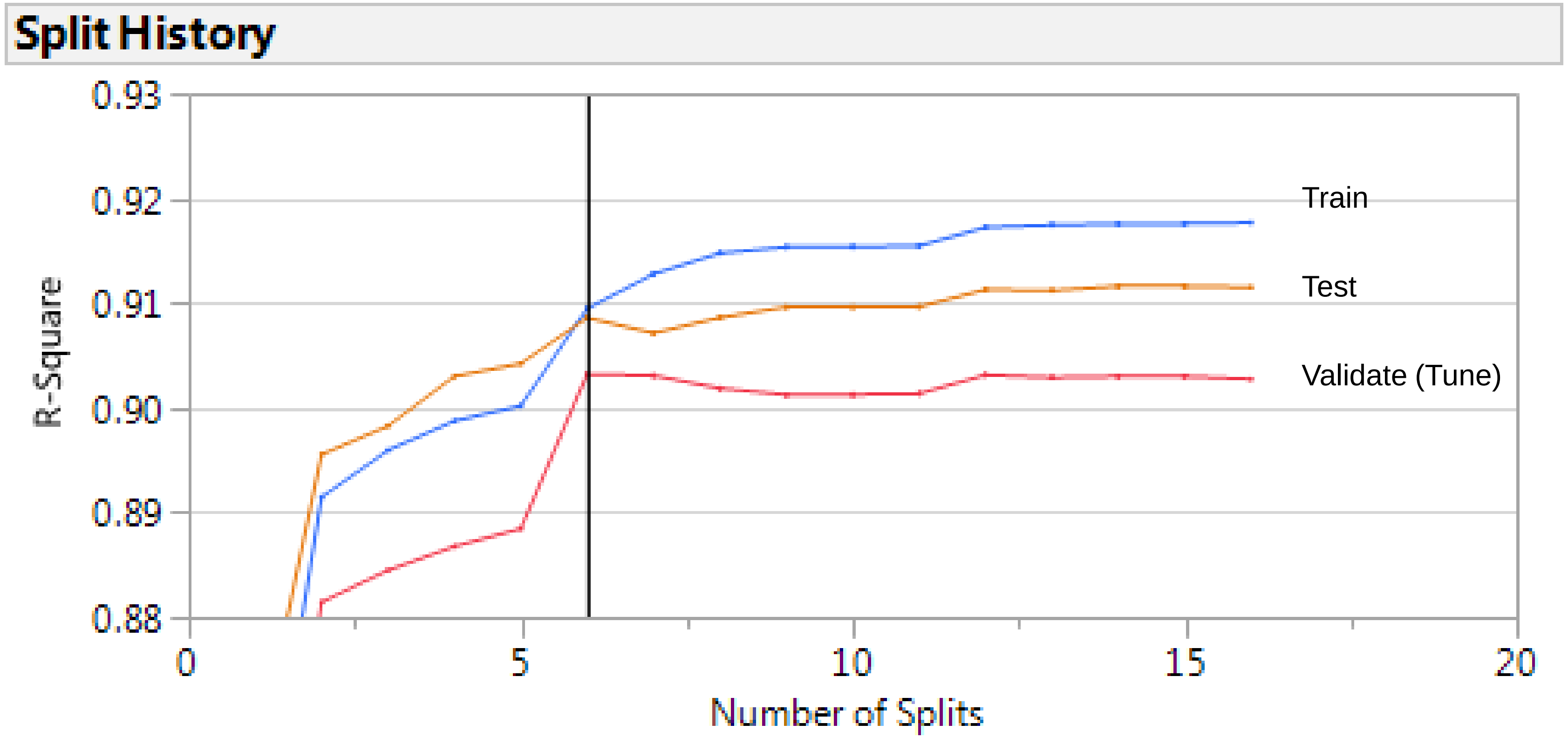


The Elements of Statistical Learning – Data Mining, Inference, and Prediction

Hastie, Tibshirani, and Friedman – 2001

(Chapter 7: Model Assessment and Selection)

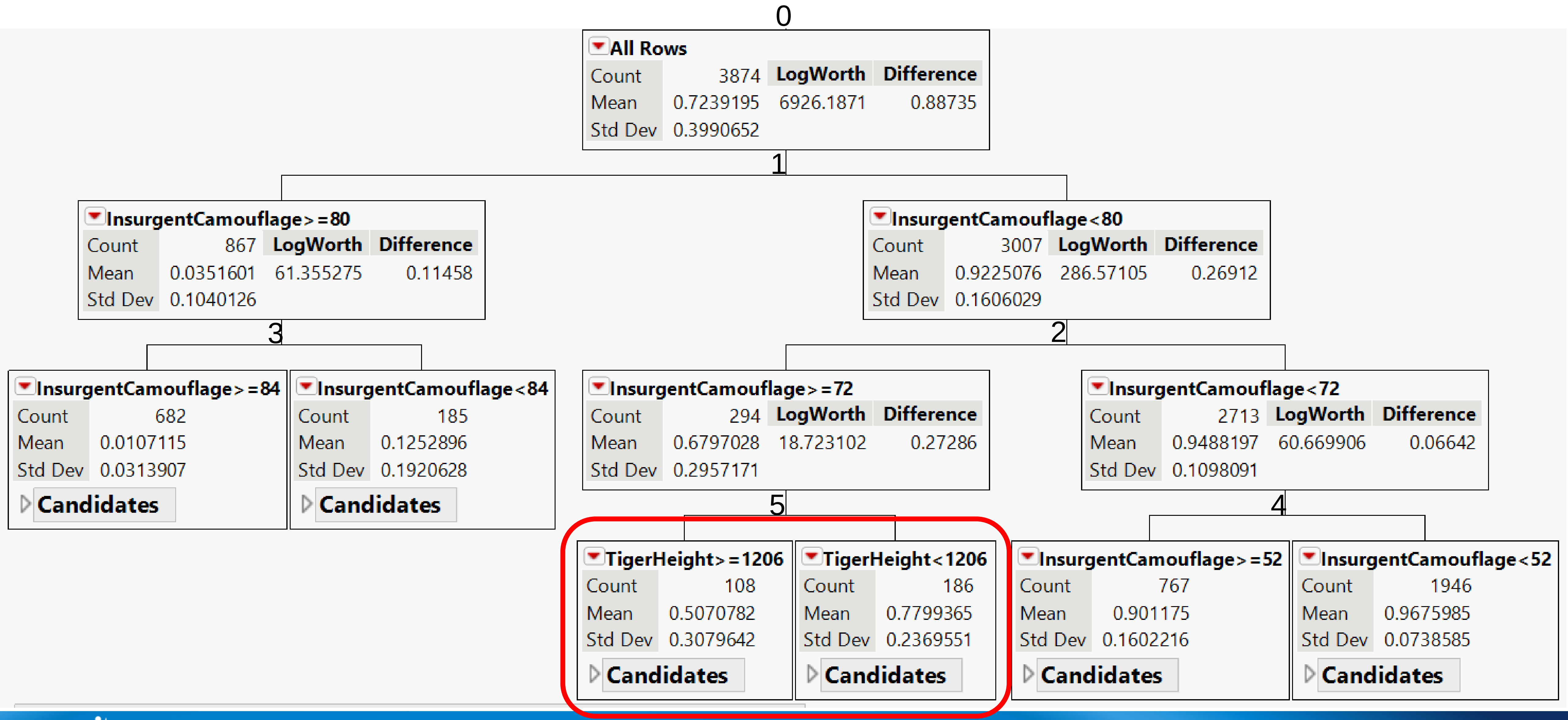
R-SQUARE VS. NUMBER OF SPLITS
(FOR A RANDOM SPLIT INTO TRAIN, VALIDATE, & TEST)



Validation Data in Red
Test Data in Orange

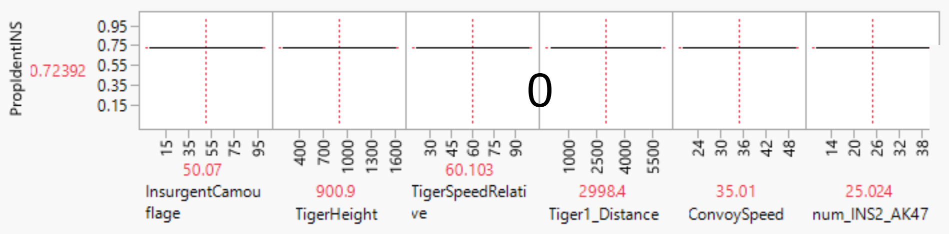
DECISION TREE

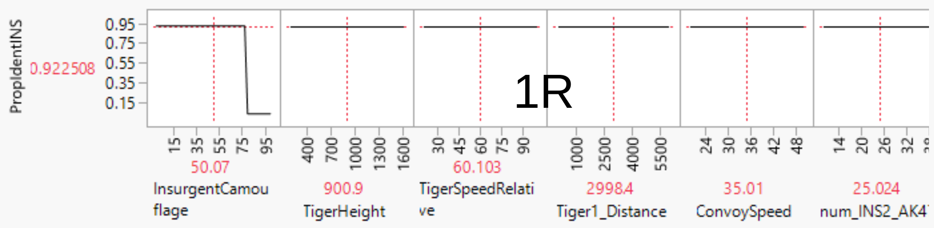
Each split finds the cut point among all factors that creates the biggest difference in the means of the two partitions of the data

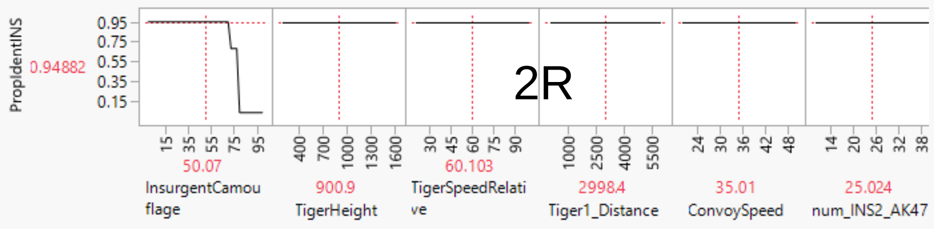


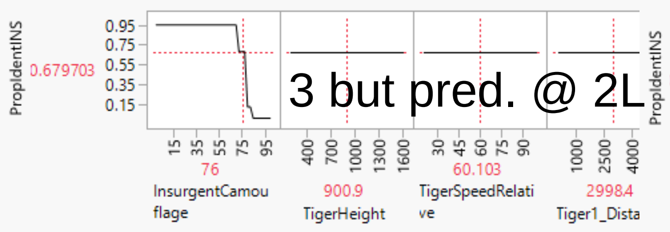
DECISION TREE

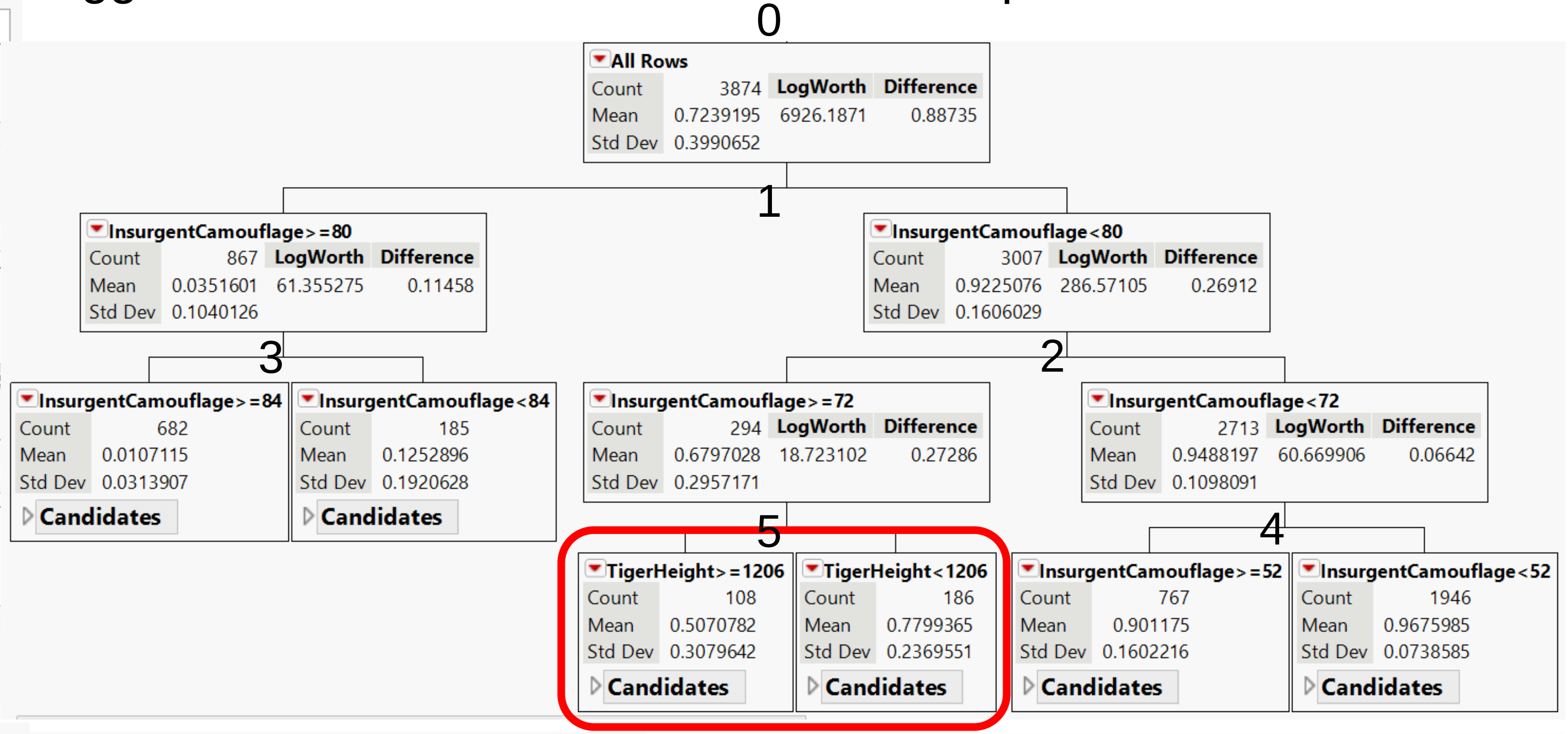
Each split finds the cut point among all factors that creates the biggest difference in the means of the two partitions of the data

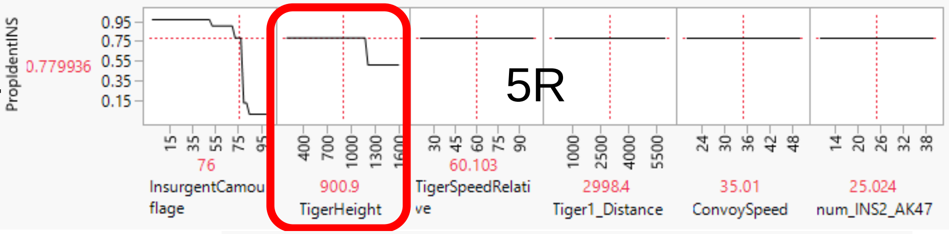


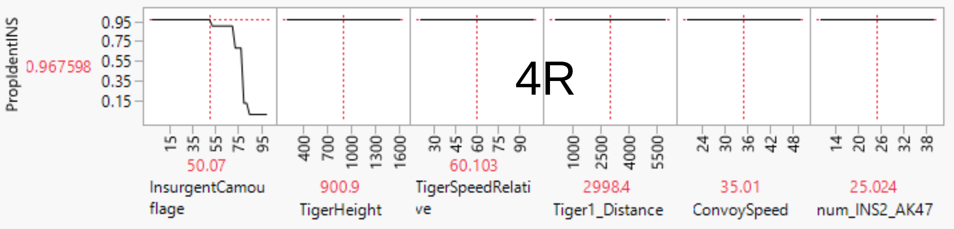




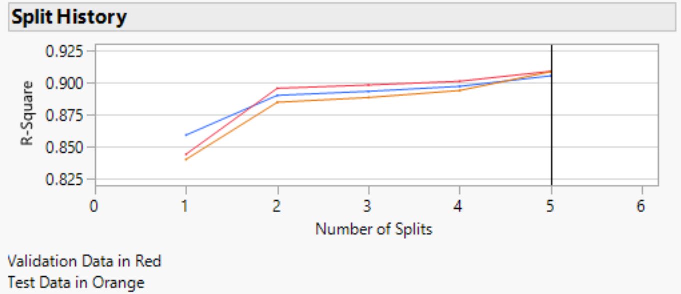








Can be interpreted as a series of nested “If” statements

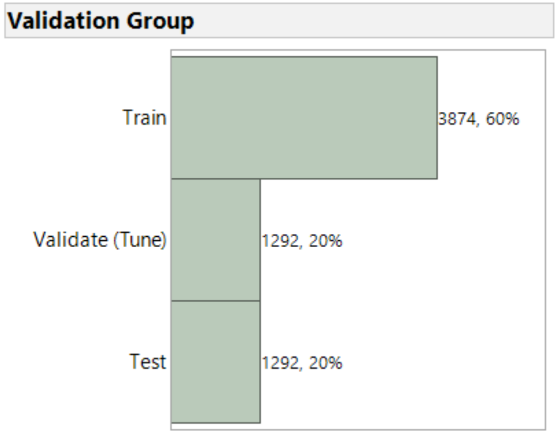


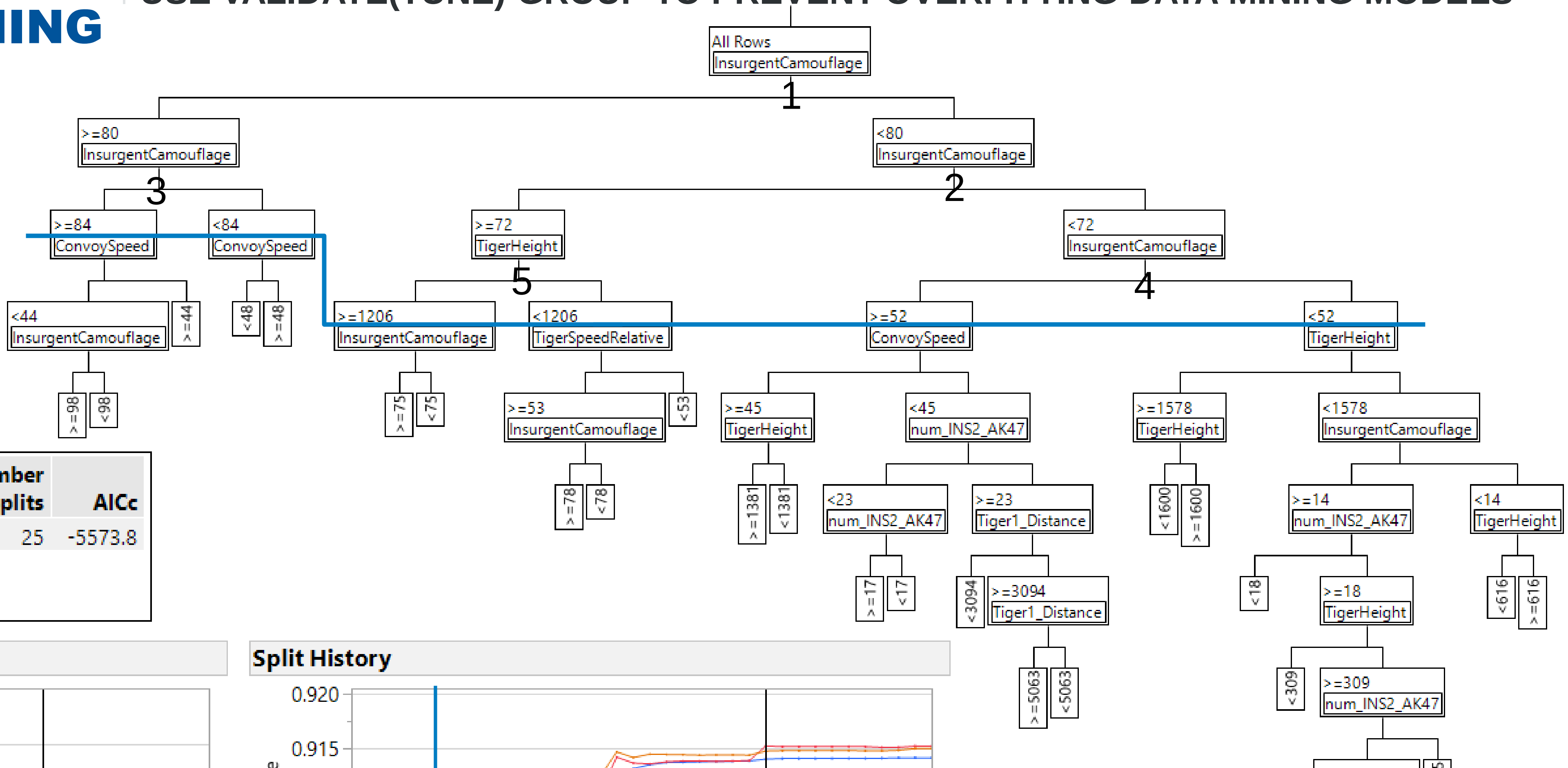
Column Contributions

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

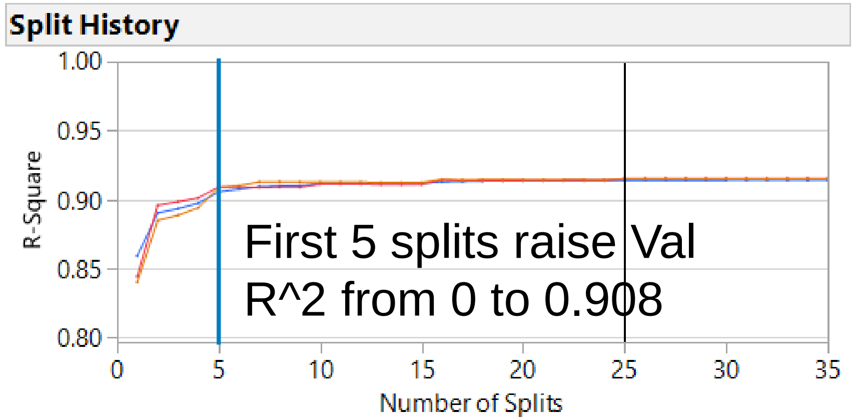
HONEST ASSESSMENT WHEN DATA MINING

SUBSET DATA TO CREATE *TRAIN*, *VALIDATE*(*TUNE*), & *TEST* GROUPS
USE VALIDATE(TUNE) GROUP TO PREVENT OVERFITTING DATA MINING MODELS

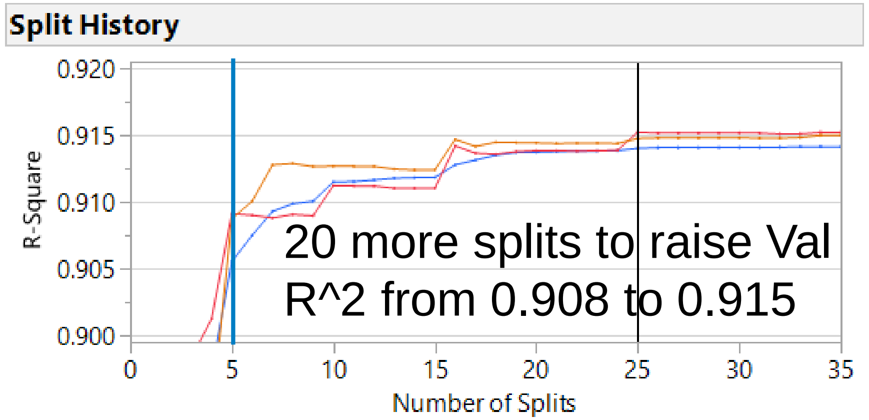




	RSquare	RMSE	N	Number of Splits	AICc
Training	0.914	0.1170276	3874	25	-5573.8
Validation	0.915	0.1132339	1292		
Test	0.915	0.1147605	1292		



Validation Data in Red
Test Data in Orange



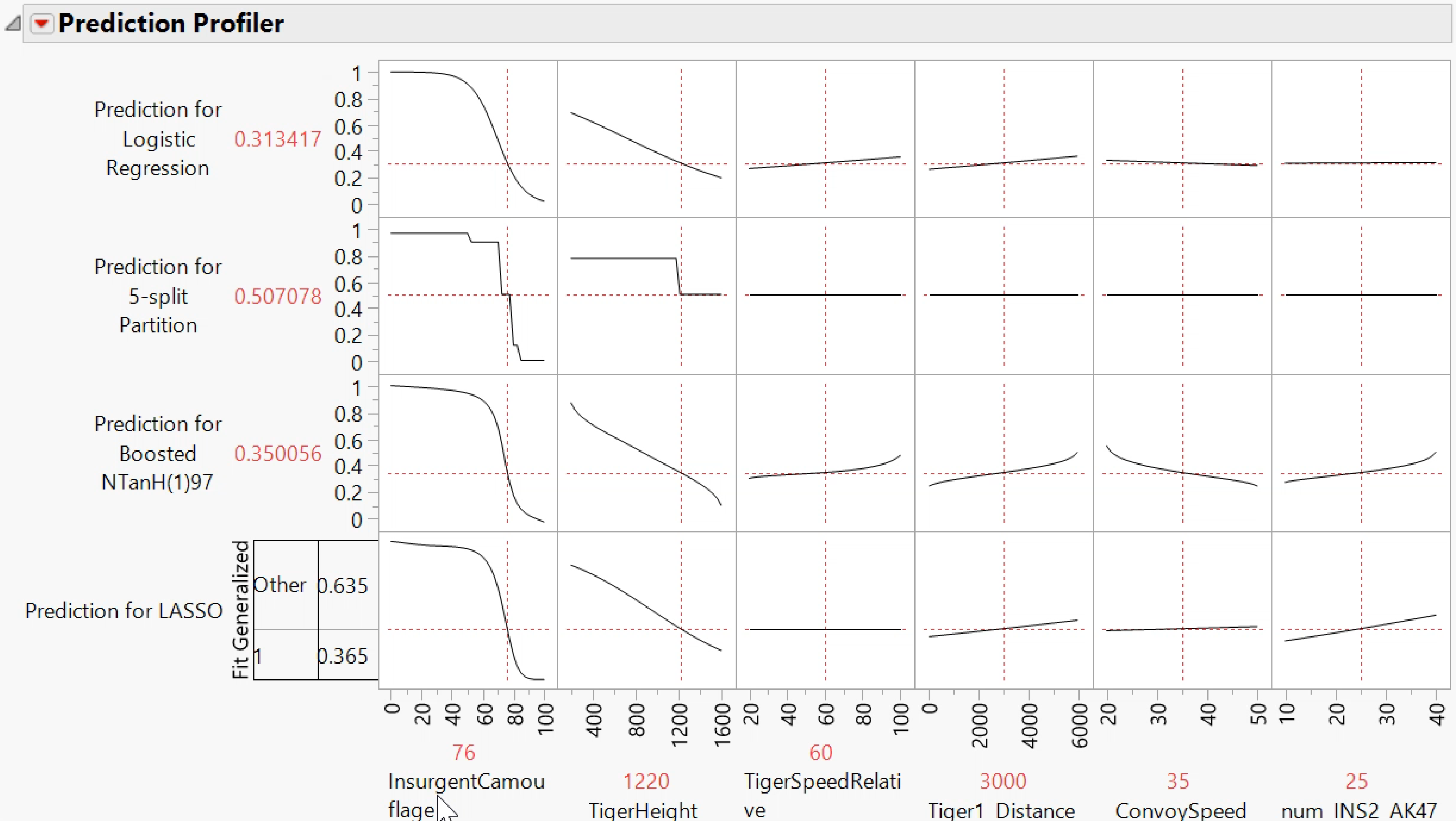
Validation Data in Red
Test Data in Orange

Column Contributions

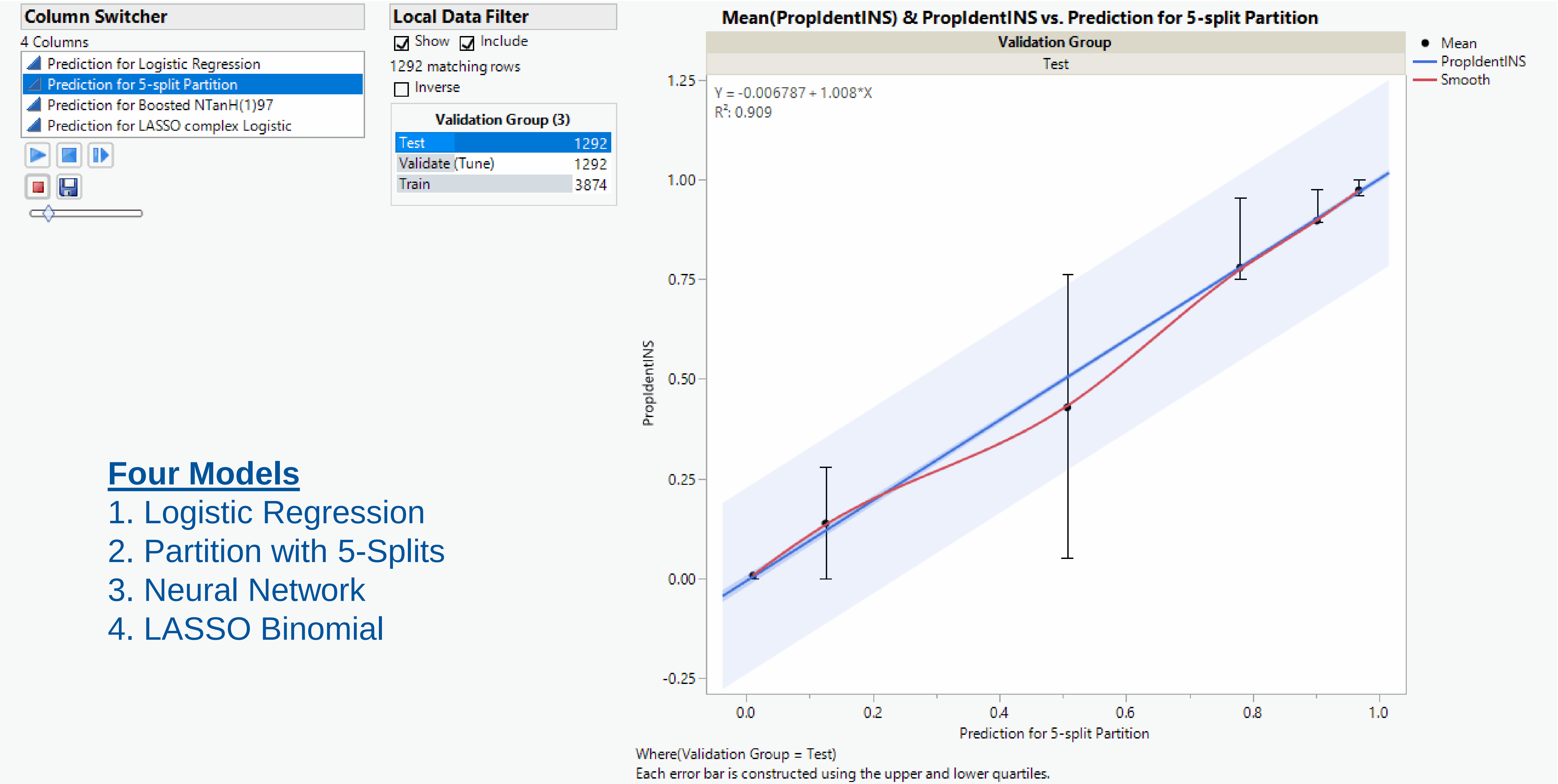
Term	Number of Splits	SS	Portion
InsurgentCamouflage	9	555.084982	0.9847
TigerHeight	6	6.46096421	0.0115
ConvoySpeed	4	1.45893941	0.0026
num_INS2_AK47	4	0.66588349	0.0012
Tiger1_Distance	2	0.06006294	0.0001
TigerSpeedRelative	0	0	0.0000

COMPARE SEVERAL MODELS

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



ACTUAL VS. PREDICTED PLOTS FOR ***TEST DATA ONLY***

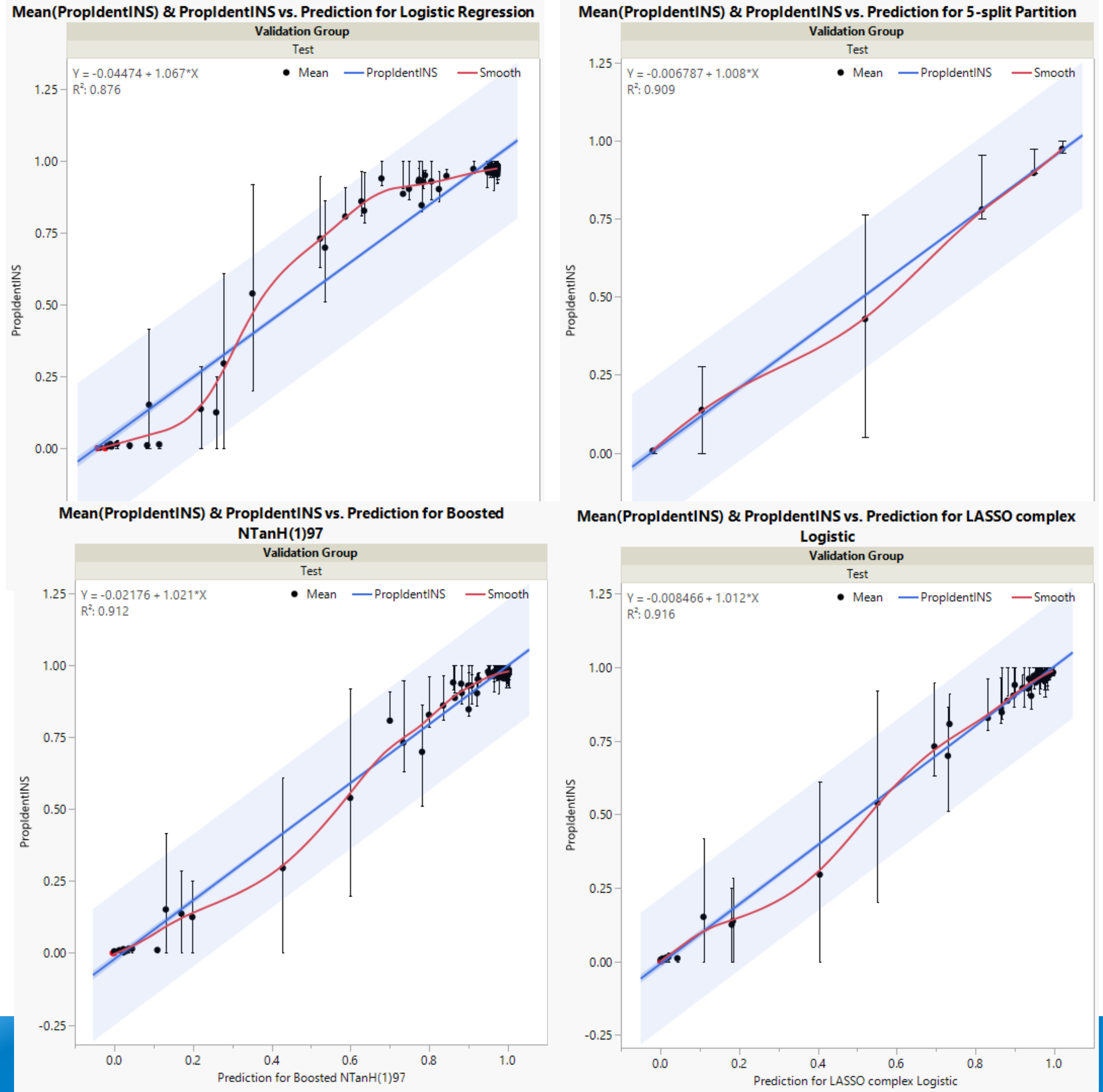


<u>Four Models</u>

1. Logistic Regression
2. Partition with 5-Splits
3. Neural Network
4. LASSO Binomial

ACTUAL VS. PREDICTED PLOTS FOR ***TEST DATA ONLY***

LOGISTIC REGRESSION
PARTITION WITH 5-SPLITS
NEURAL NETWORK
LASSO BINOMIAL



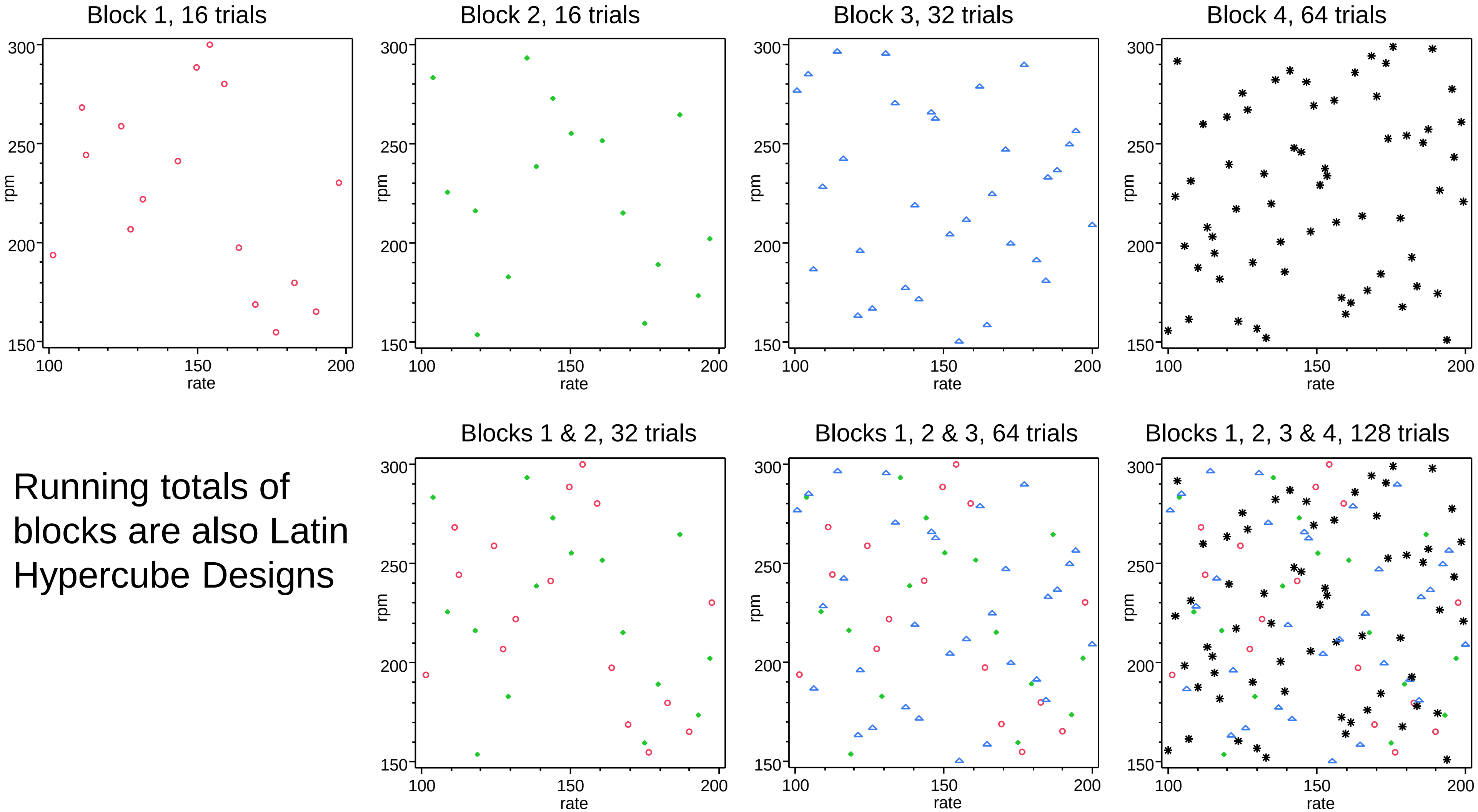
WHY IS A SEQUENTIAL APPROACH SO USEFUL?

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 50 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

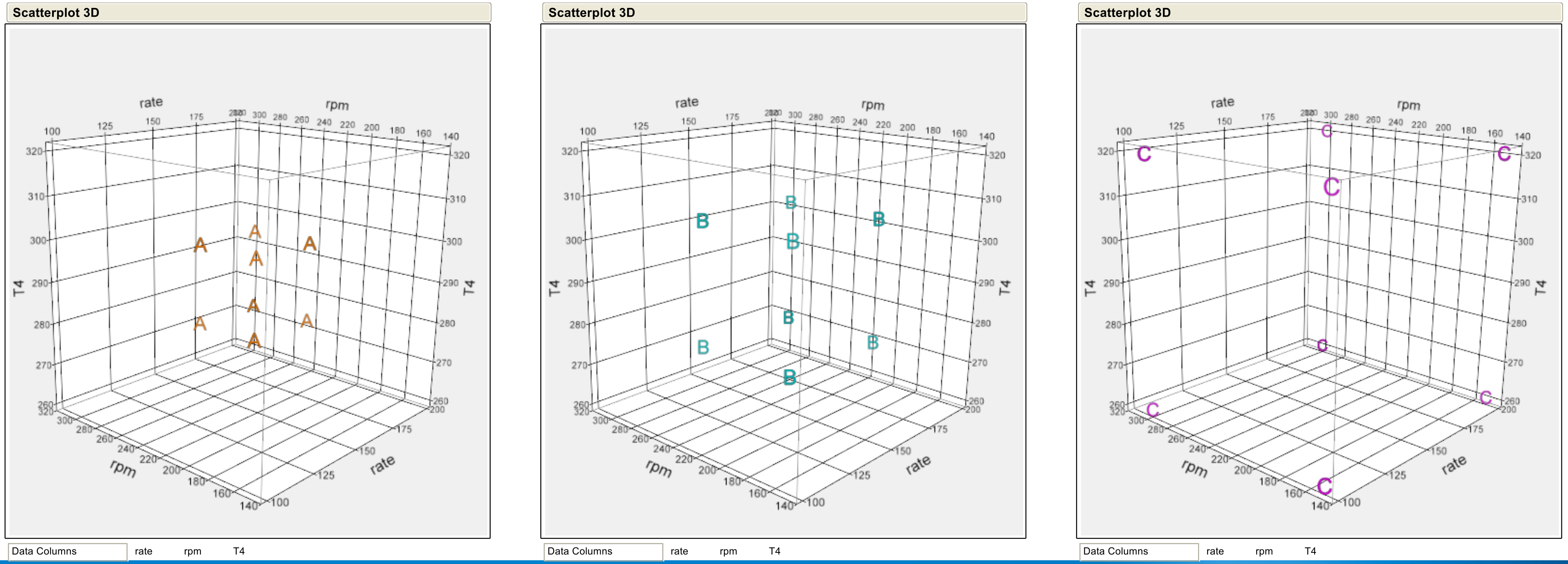


Running totals of blocks are also Latin Hypercube Designs

WHY RUN SIMULATIONS IN SEQUENTIAL BLOCKS?

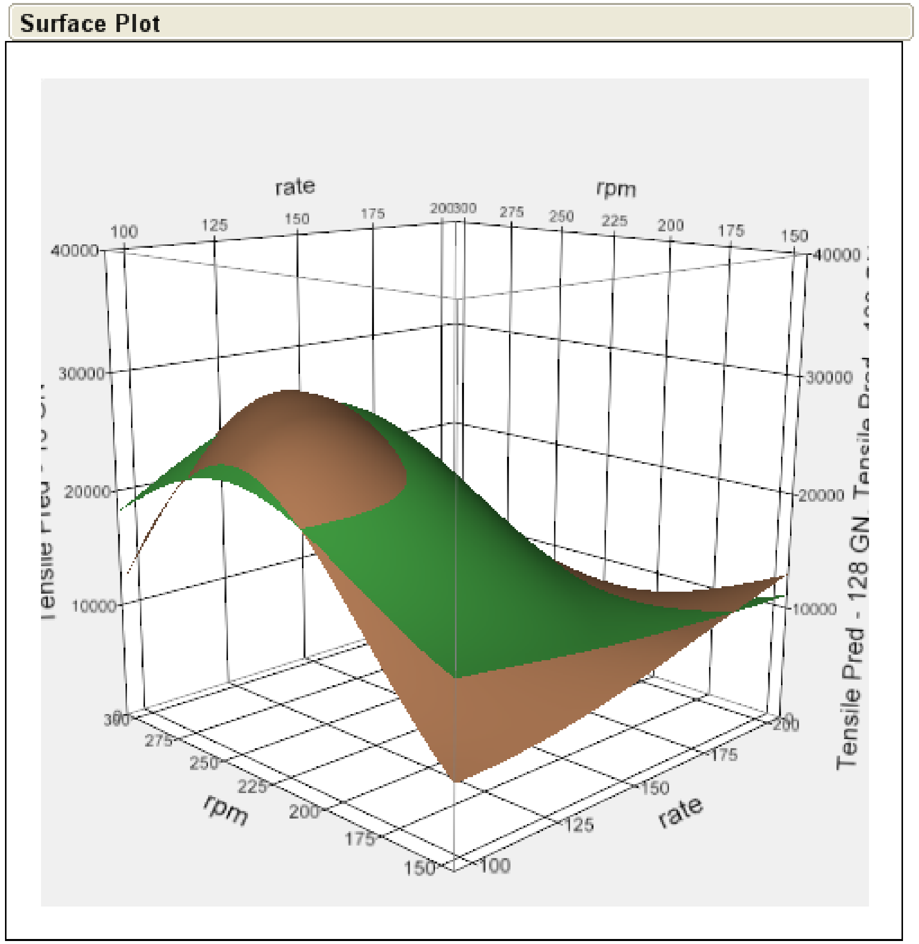
The point of running this sequence of blocks is to be able to evaluate the surrogate model after each stage to see how accurately it is predicting observed values of 3 sets of checkpoint trials. If it proves to be sufficiently accurate, then subsequent blocks of simulation trials need not be run.

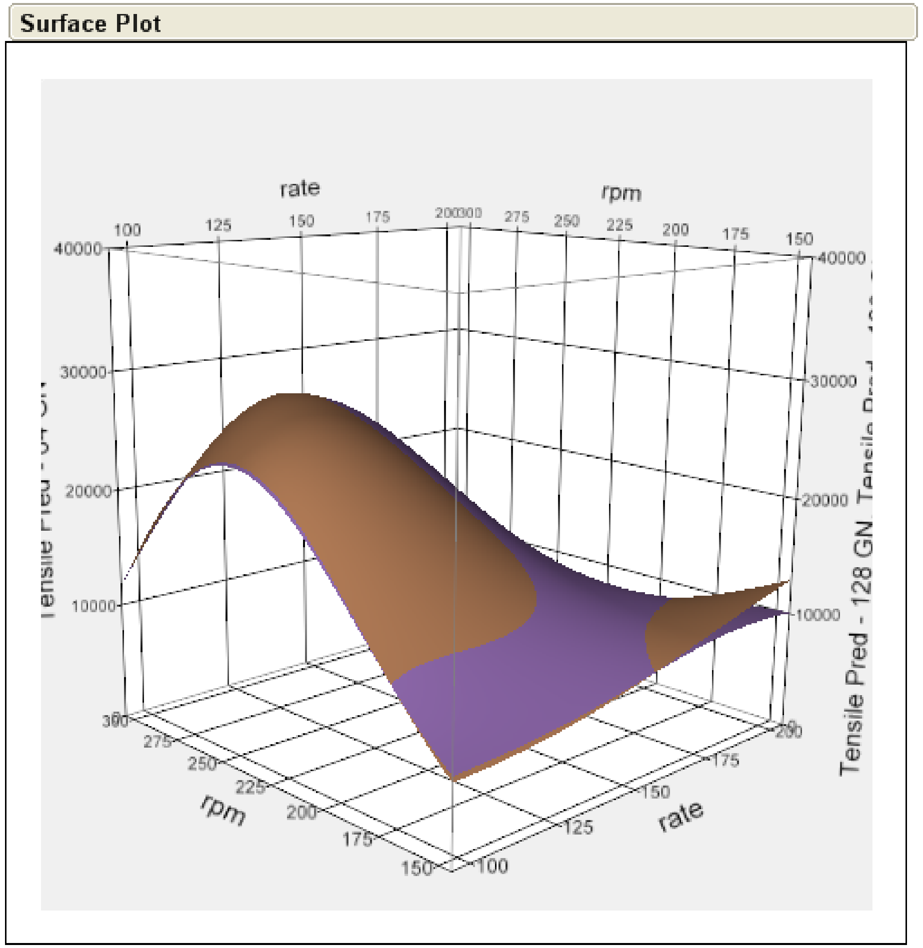
Without the NLHD approach one has to choose the "right" size space-filling design in order to get useful results. If you choose too small a design, one has to start over with a larger design.



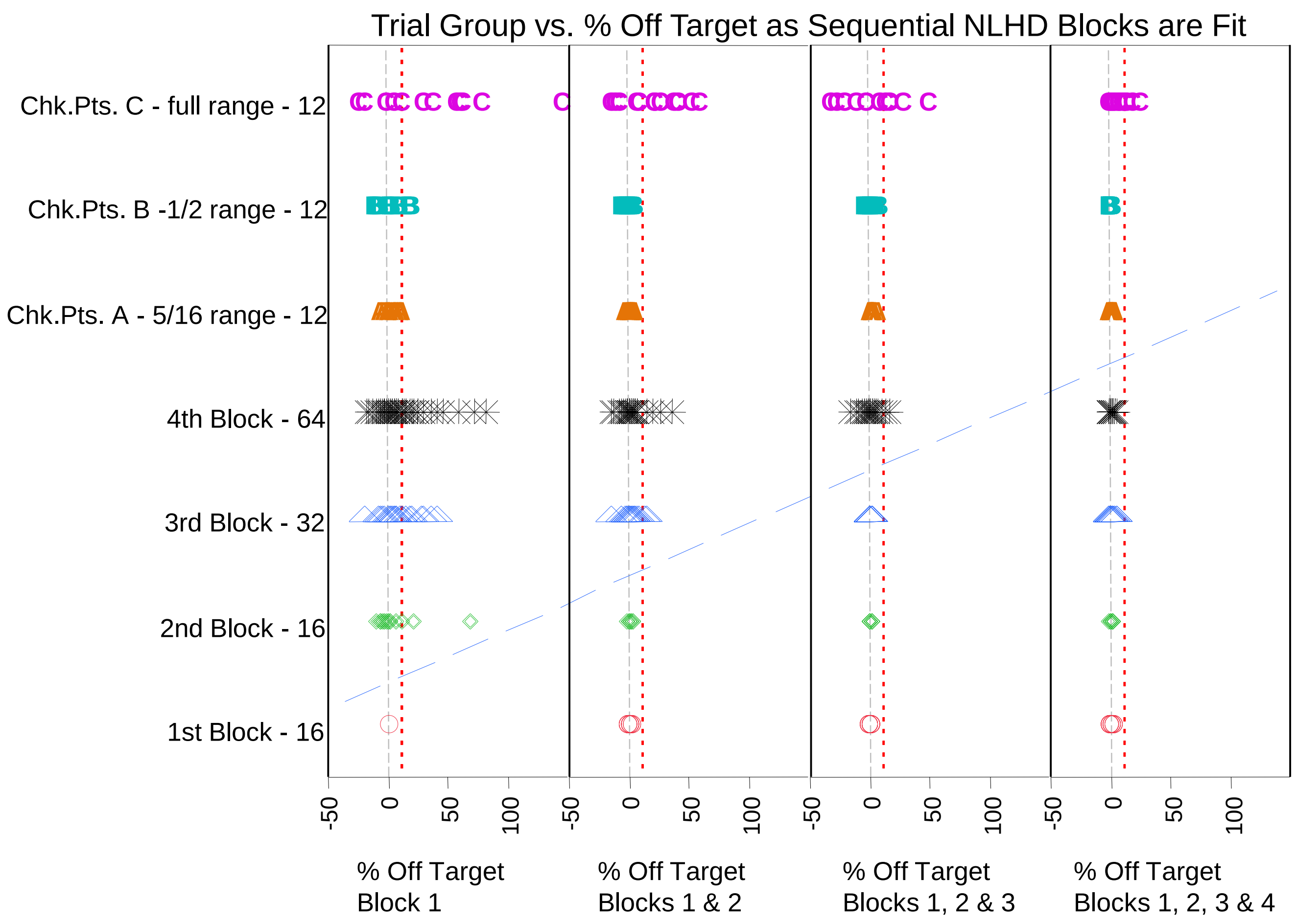
COMPARE RESPONSE SURFACES FOR FIT OF 16 VS. FIT OF 128 TRIALS (LEFT) AND FOR FIT OF 64 VS. FIT OF 128 TRIALS (RIGHT)

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





ACCURACY OF SURROGATE PREDICTIONS FOR 3 GROUPS OF CHECK-POINTS YIELDING MARGINAL, MODERATE AND EXTREME EXTRAPOLATION

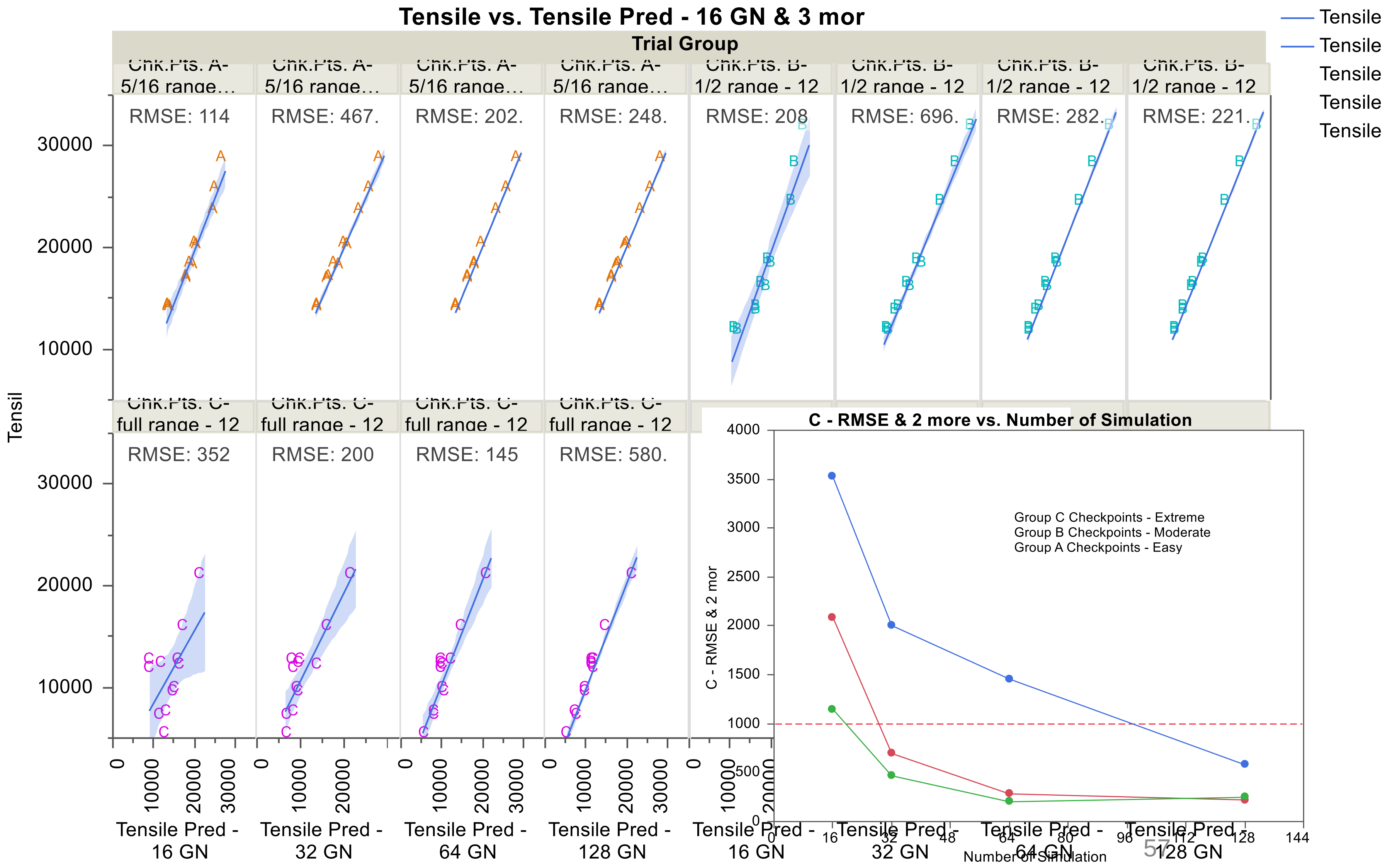


% Off Target for Checkpoints NOT Included in Model Fit

% 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



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..

WHY USE DESIGN OF EXPERIMENTS METHODS WITH SIMULATION EXPERIMENTS?

Quicker answers, lower costs*, solve bigger problems*

- Obtain a fast surrogate model of the 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 yields 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, one can – for the same resources and constraints – ***solve bigger problems***