Design of Experiments (DOE) for Real-World Problems

NDIA 26th Annual
National Test & Evaluation Conference
March 2, 2010

Tom Donnelly, PhD
JMP Principal Customer Advocate
and Co-Insurrectionist

tom.donnelly@jmp.com

Statistical Discovery. From SAS®

My Background

Design of Experiments (DOE) for 25+ Years

- ‘83-’87 Honeywell, Inc., Engineer

  First saw the power of DOE in 1984 – career changing event
- ‘87-’99 ECHIP, Inc., Partner & Technical Director

  200+ DOE courses, on-site at 40+ companies - many chemical/food/pharma - requiring mixture/formulation DOE
- ‘99-’05 Peak Process, LLC, Consultant
- ‘05-’08 US Army, Edgewood CB Center, Analyst

  DOE with Real data and Modeling & Simulation data
- Dec. ’08 Joined the SAS Institute Inc., Customer Advocate

  Work in DOE and Federal Government domains
  - Data Visualization, Data Mining and their synergy with DOE
  - Primarily support DoD sites and National Laboratories

Projects Using DOE at U.S. Army ECBC

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 experimental designs used with simulations
  - Re-analysis of U.S. Air Force Kunsan 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

Key Take Away Points

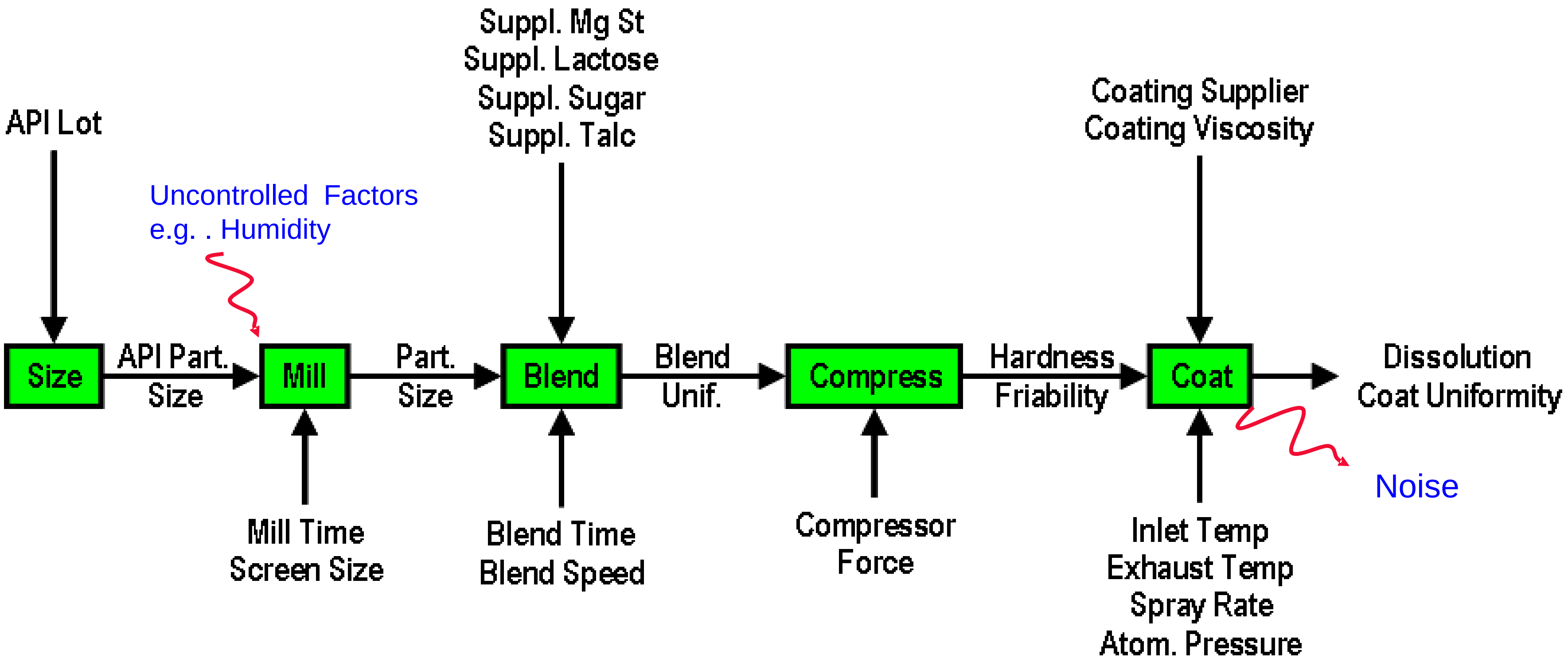
- Building predictive models of multiple responses provides decision makers with the knowledge to make better informed judgments and trade-offs.
- A Design of Experiments (DOE) is a collection of trials built to support a proposed model.
- An algorithmic design tool can quickly build a DOE for your predictive model
  - subject to real-world physical constraints
  - sequence of DOEs can be built to integrate testing
- Tools are great, but education is more important!
  "Successful use of DOE will require a cadre of personnel… with the professional knowledge and expertise…"

Three sections to this part of the tutorial

- **Overview and alternative way to think about Design of Experiments (DOE)**
- The power of predictive modeling
  - Show how you can efficiently deliver process knowledge that makes the jobs of decision makers easier
- Example real-world DOEs
  - Quickly create a design for a proposed model subject to real-world physical constraints
  - Generate a sequence of designs for increasingly complex models and to integrate testing

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).



Here are 4 Controls (inputs) & 2 Responses (outputs) and their empirical relationships (model)

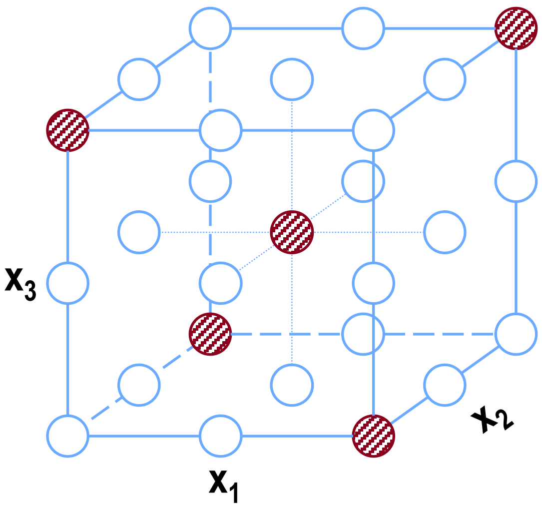
Get this Prediction Profiler as result of analyzing data collected for a DOE

Alternative Definition

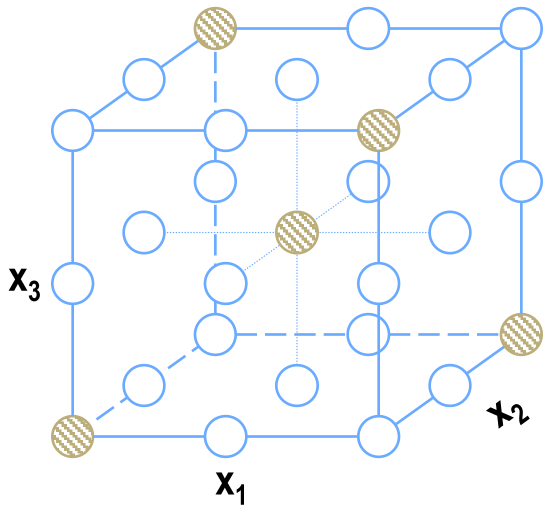
- A DOE is the specific collection of trials run to support a proposed model.
  - If proposed model is ***simple***, e.g. just main or 1st order effects (x_1 , x_2 , x_3, etc.), the design is called a ***screening*** DOE
    - Goals include **rank factor importance** or find a “winner” quickly
    - Used with many (> 6?) factors at start of process characterization
  - If the proposed model is ***more complex***, e.g. the model is 2nd order so that it includes two-way interaction terms (x_1x_2 , x_1x_3, x_2x_3, etc.) and in the case of continuous factors, squared terms (x_1^2, x_2^2, x_3^2 , etc.), the design is called a ***response-surface*** DOE
    - Goal is generally to develop a **predictive model** of the process
    - Used with a few (< 6?) factors after a screening DOE

Classic Response-Surface DOE in a Nutshell

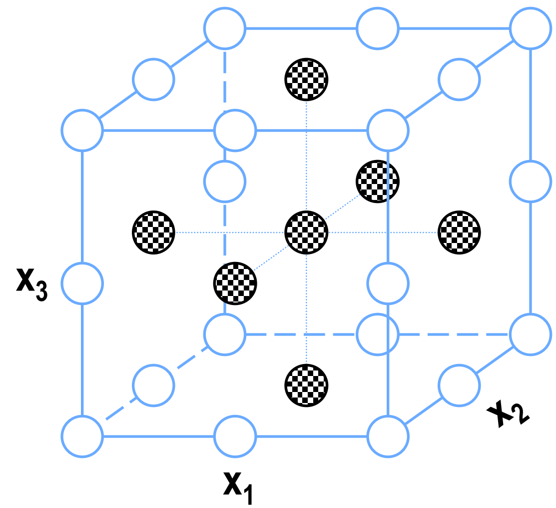
Block 1



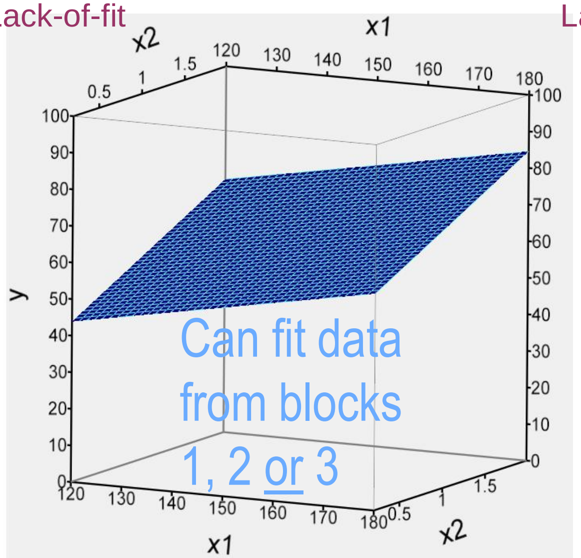
Block 2



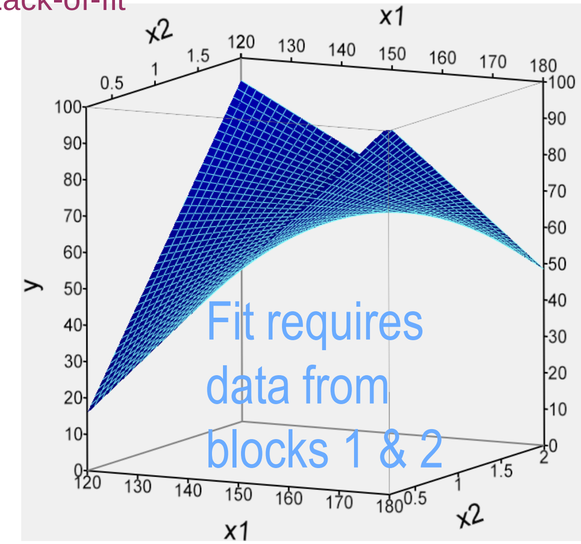
Block 3

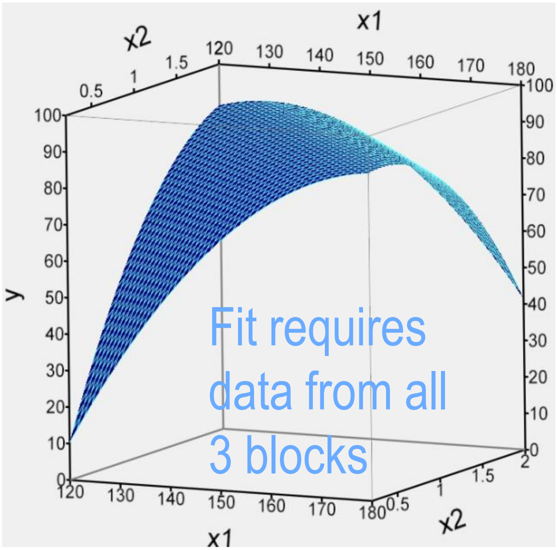


Lack-of-fit



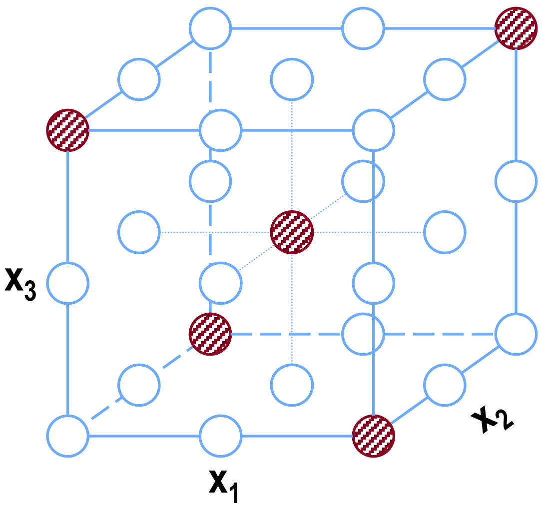
Lack-of-fit





Expensive Experimentation? Sequential DOE Can Be Used

Block 1



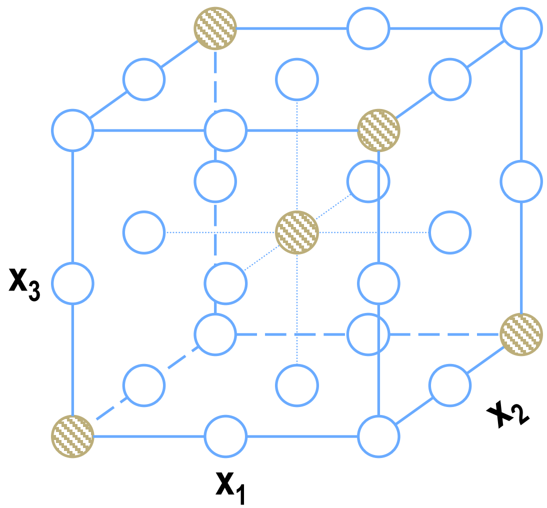
$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$

Run this block 1st to:

(i) estimate the main effects*
(ii) use center point to check for curvature.

*May be all that are needed with appropriate physics-based scaling
Also called "non-linear modeling"

Block 2



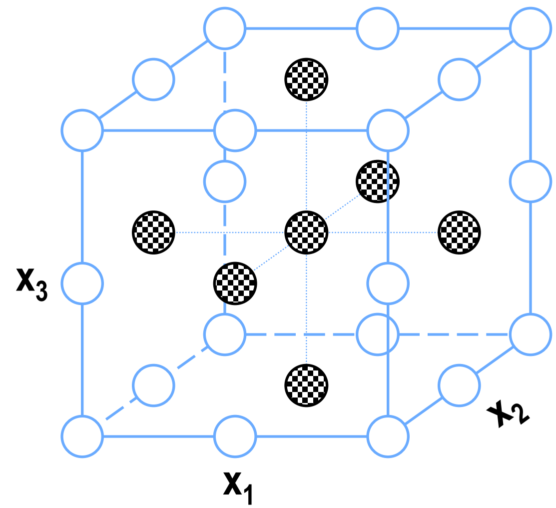
$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$

$+ a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3$

Run this block 2nd to:

(i) repeat main effects estimate,
(ii) check if process has shifted
(iii) add interaction effects to model <u>if needed.</u>

Block 3



$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3$

$+ a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3$

$+ a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2$

Run this block 3rd to:

(i) repeat main effects estimate,
(ii) check if process has shifted
(iii) add curvature effects to model <u>if needed.</u>

Why Use Design of Experiments (DOE)?

Quicker answers, lower costs, solve bigger problems to better protect the warfighter

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."*

Robert M. Gates, Secretary of Defense, January 27, 2009

Why Use Design of Experiments (DOE)?

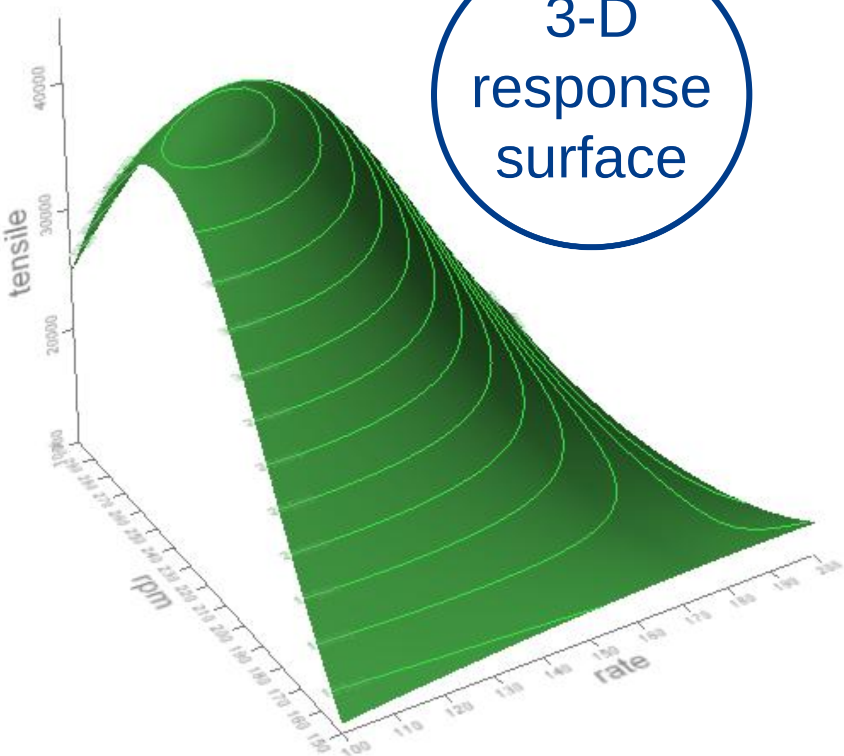
Quicker answers, lower costs, solve bigger problems

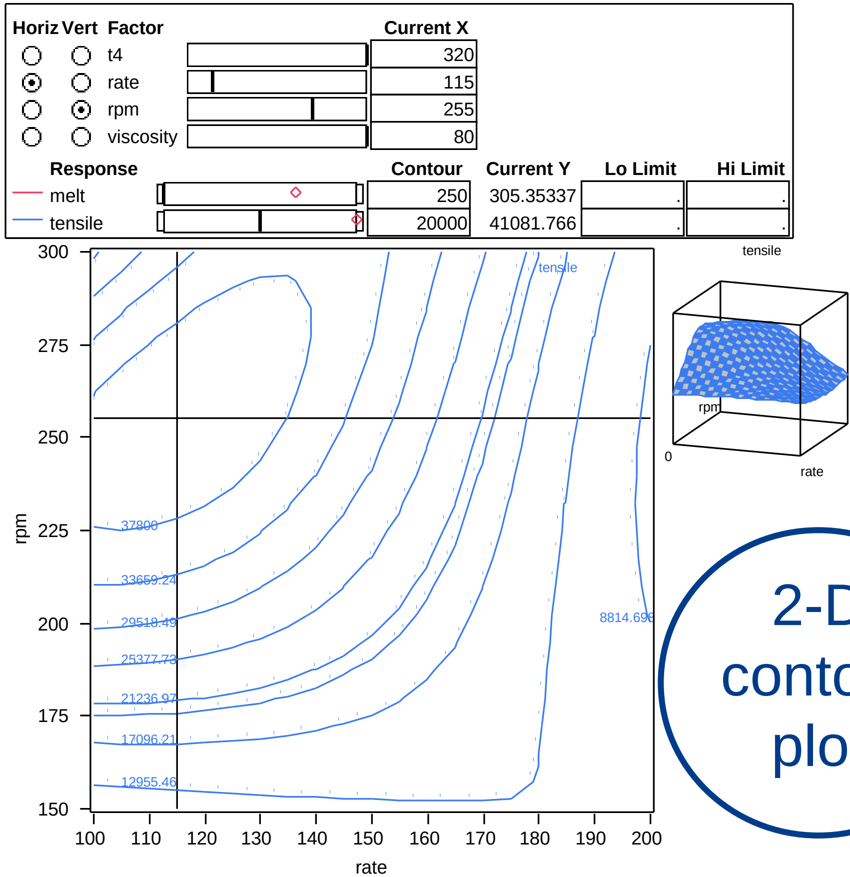
- Real Data
  - Get a ranking of the factors – pick a winner
  - Get a predictive “picture” (with 95% limits) of the process
- Simulation Data – used more and more in DoD and Industry
  - Obtain a fast surrogate “metamodel” of the long-running simulation
- Analysis benefits for both types of data:
  - more rapidly answer “what if?” questions
  - do sensitivity analysis of the control factors
  - optimize multiple responses and make trade-offs
- By running efficient subsets of all possible combinations, one can – for the same resources and constraints – *solve bigger problems*
- 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

Response Surface & Contour Plot

(four control variables)

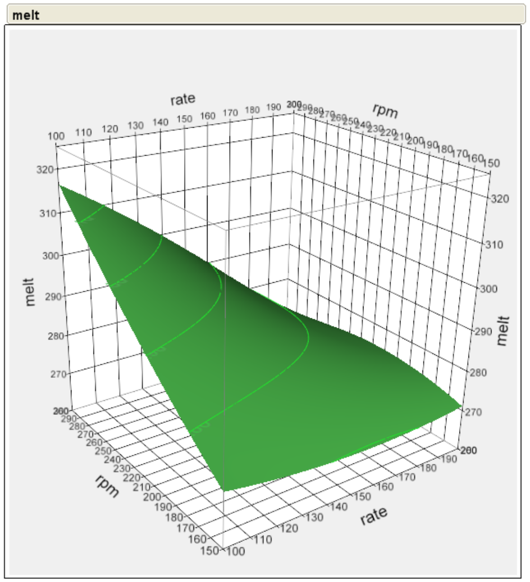
3-D response surface

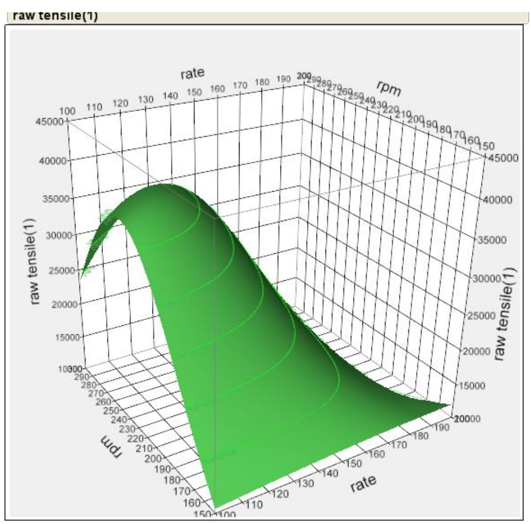




2-D contour plot

Response Surfaces & Contour Plots
(two responses and four control variables)

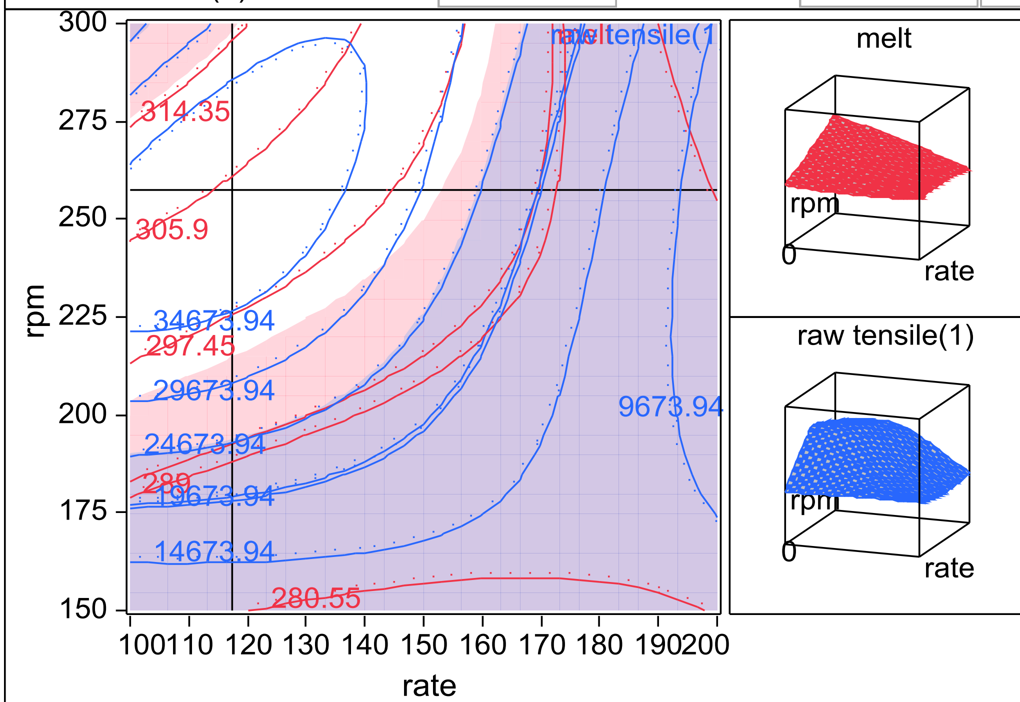




Contour Profiler

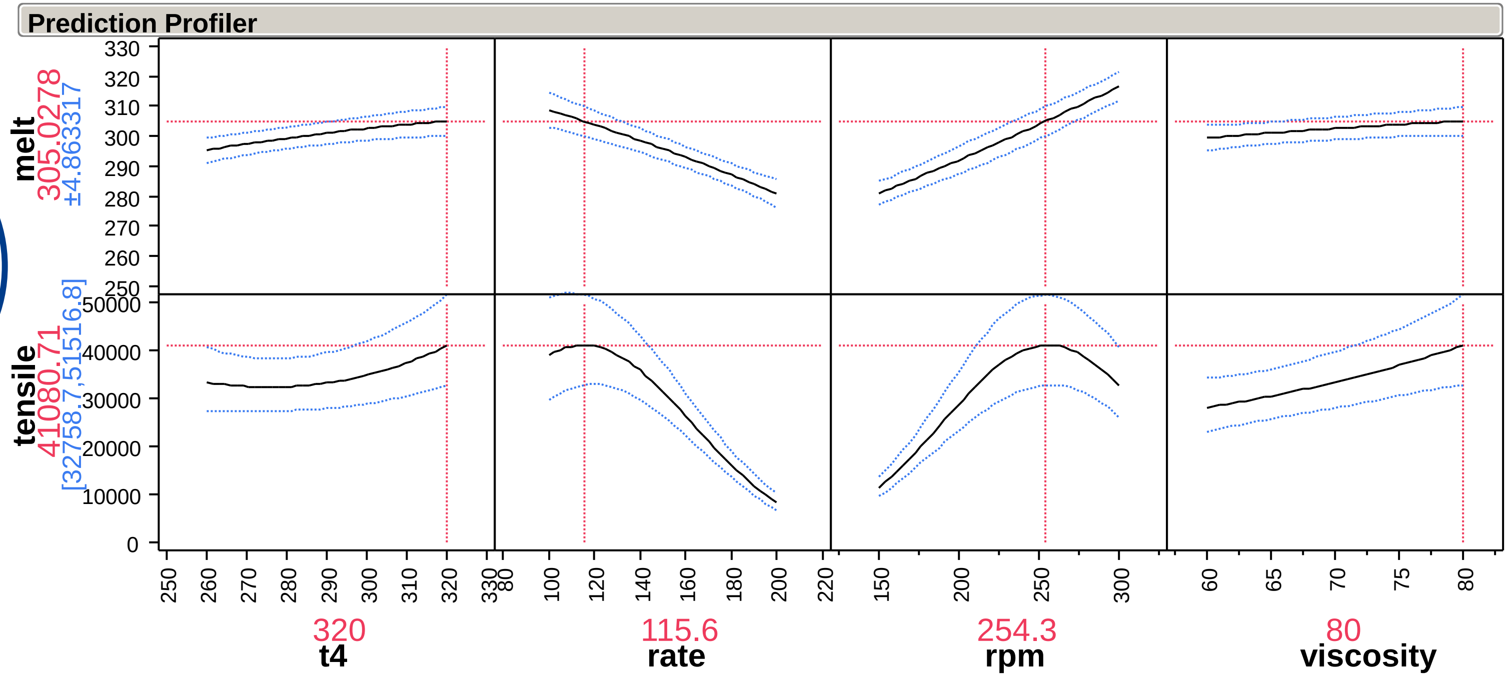
HorizVert	Factor	Current X
	t4	320
	rate	117.29498
	rpm	257.64505
	viscosity	80

Response	Contour	Current Y	Lo Limit	Hi Limit
melt	290	305.06654	295	315
raw tensile(1)	20000	38127.616	25000	40000

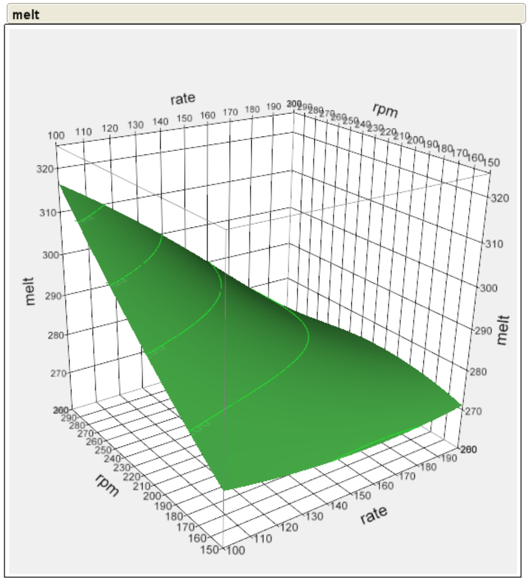


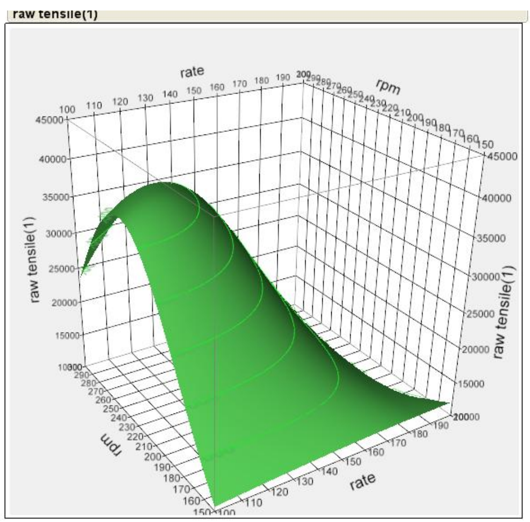
1-D Prediction Profiles are a Way to View Higher Dimensionality as "Interactive Small Multiples" - Here 4 Controls & 2 Responses

1-D profiler plots



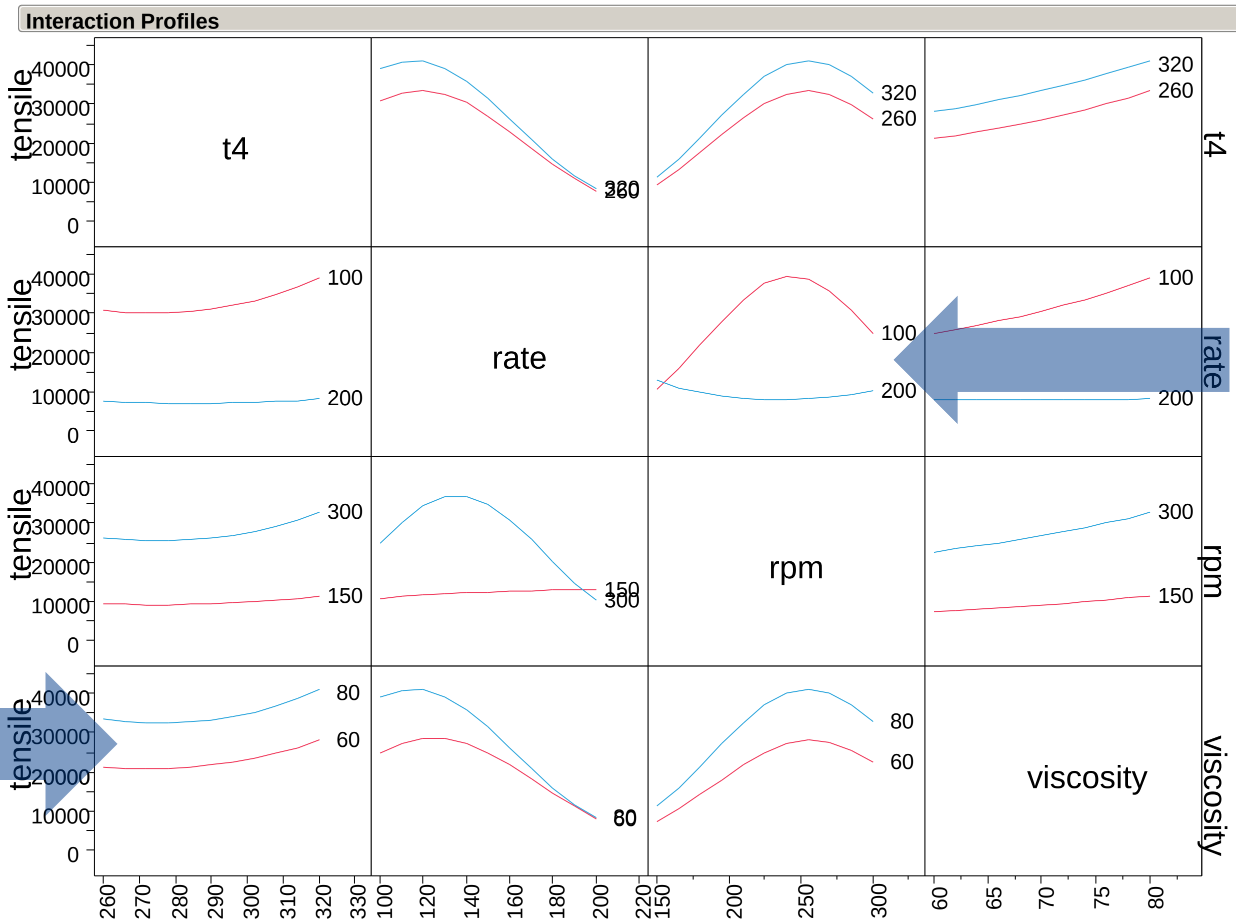
1-D Prediction Profiles are a Way to View Higher Dimensionality as "Interactive Small Multiples" - Here 4 Controls & 2 Responses





Interaction Profiles are Another Way to View Higher Dimensionality - Here 4 Controls and 1 Response

1-D plots at high & low of other factors



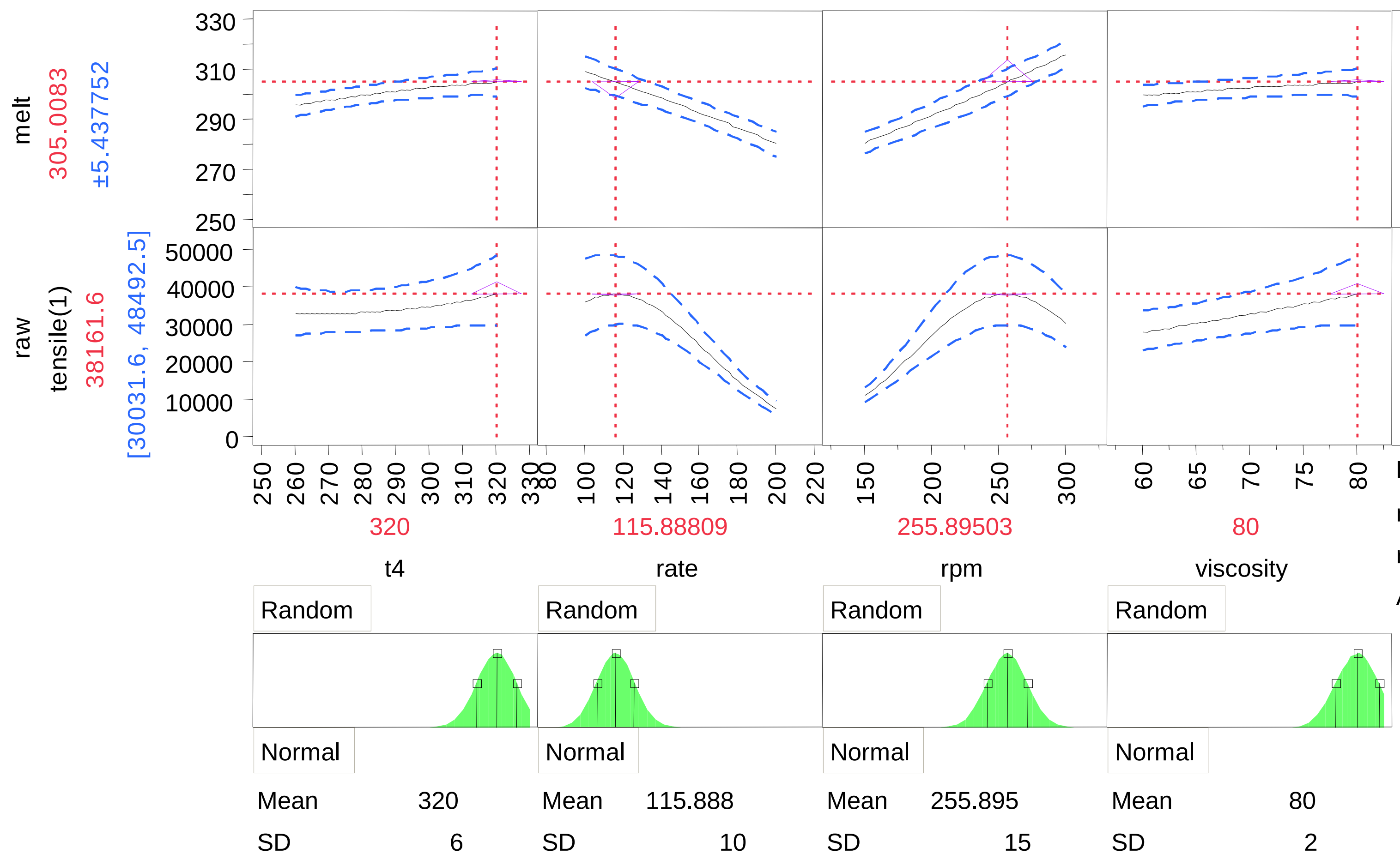
NOT Parallel indicates interaction

Parallel indicates NO interaction

Find Robust Operating Conditions that are Insensitive to Variability in Control Factors

Monte Carlo simulations can be run using known or assumed distributions of input variability to better assess transmitted variation about the model point estimate.

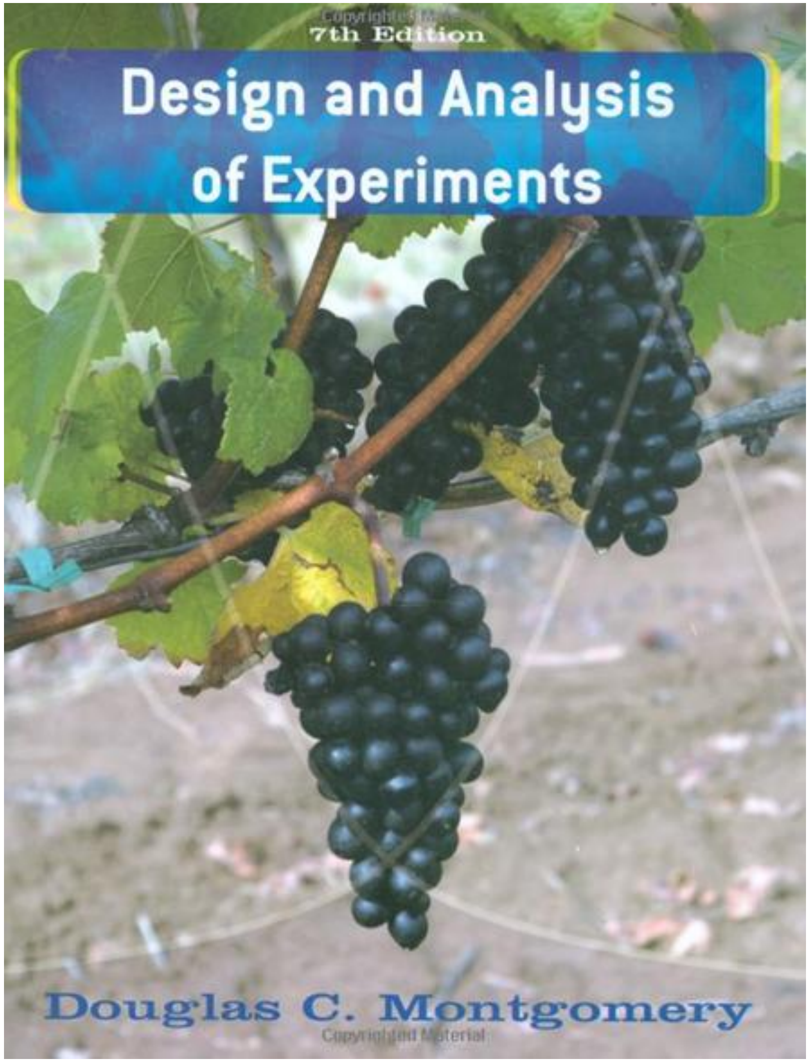
Prediction Profiler



Defect	Rate
melt	0.0316
raw tensile(1)	0.0024
All	0.0316

Mean	SD
304.959	4.62913
36490.9	2566.08

DOE Textbook (and Companion Student Guide) with Examples Using Leading Software Packages



Many other resources listed on MORS Symposium Real-World DOE Tutorial PDF
MORS Symposium is June 21-24, 2010, Quantico, VA www.mors.org

Three sections to this part of the tutorial

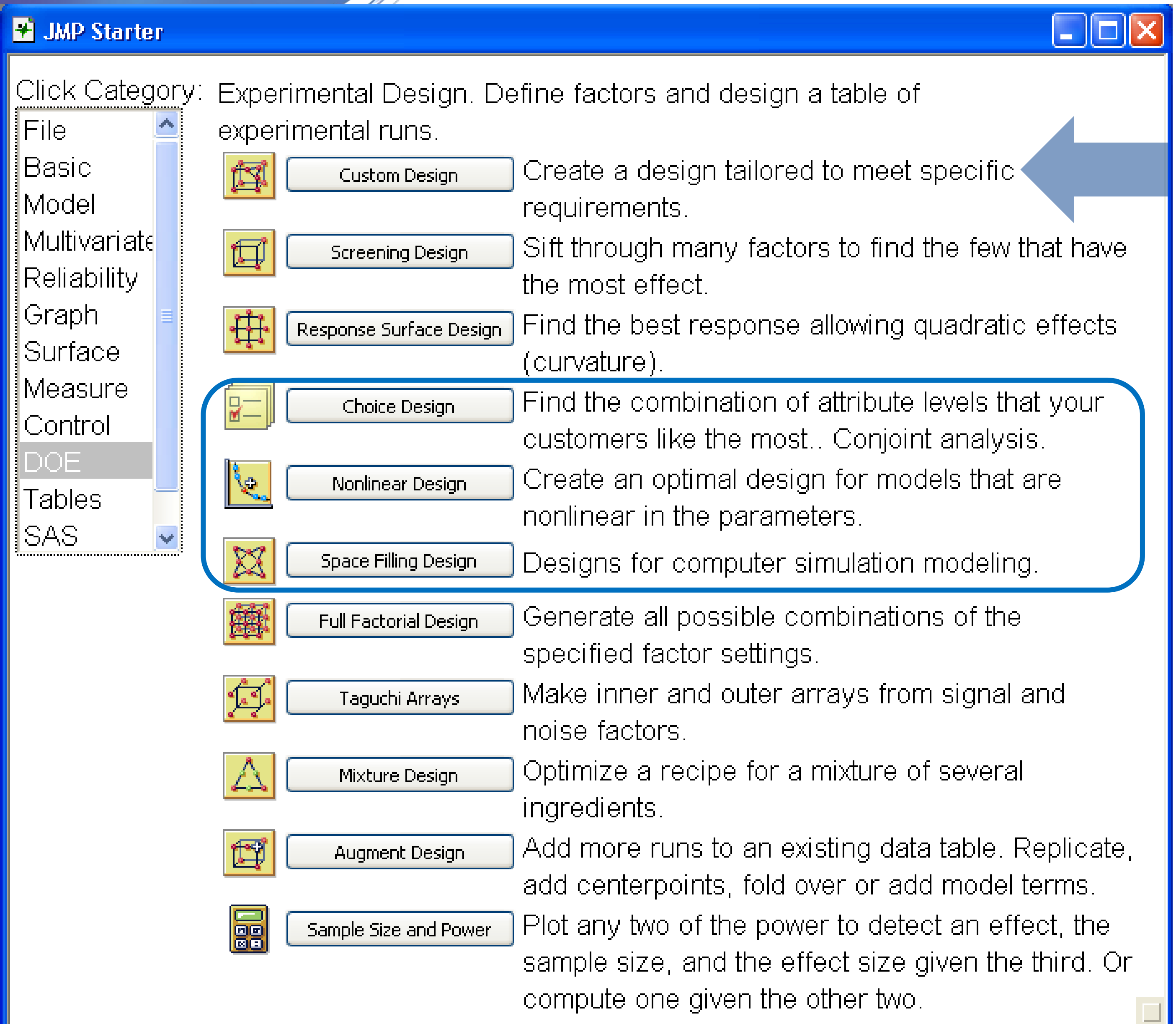
- Alternative way to think about making a Design of Experiments (DOE)
- **The power of predictive modeling**
  - Show how you can efficiently deliver process knowledge that makes the jobs of decision makers easier
- Example real-world DOEs
  - Quickly create a design for a proposed model subject to real-world physical constraints
  - Generate a sequence of designs for increasingly complex models and to integrate testing

Multiple Response Optimization
3 responses and 4 control factors

Three sections to this part of the tutorial

- Alternative way to think about making a Design of Experiments (DOE)
- The power of predictive modeling
  - Show how you can efficiently deliver process knowledge that makes the jobs of decision makers easier
- **Example real-world DOEs**
  - Quickly create a design for a proposed model subject to real-world physical constraints
  - Generate a sequence of designs for increasingly complex models and to integrate testing

Many Types of Designs for Many Types of Problems



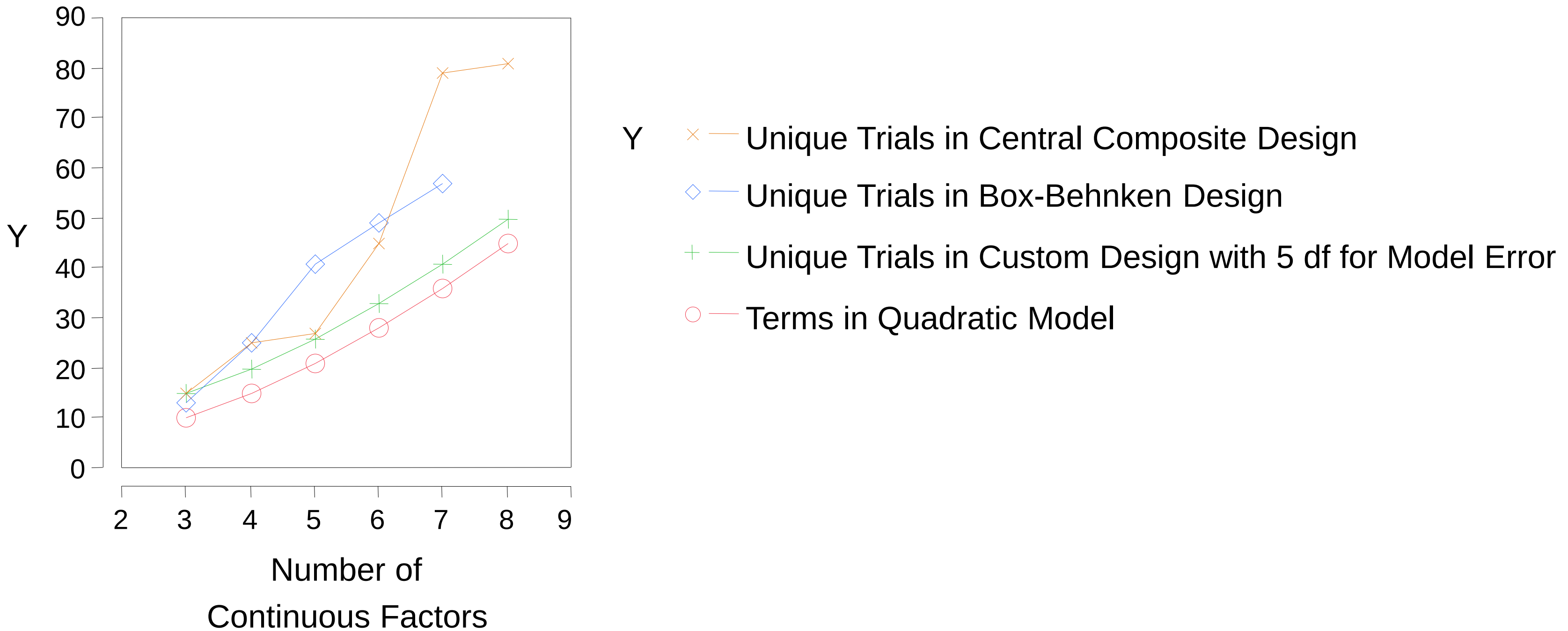
The “real-world” DOE solution that’s good to use even when the problem is simple.

Design methods for very specialized problem areas.

Create a Straightforward Custom DOE

- Enter Factor and Response Information
  - Responses – *Speed*, *Contrast* and *Cost*
  - Factors and ranges (or levels):
    - *Sensitizer 1* 50 to 90
    - *Sensitizer 2* 50 to 90
    - *Dye* 200 to 300
    - *Reaction Time* 120 to 180
- Propose Model
  - 2^{nd} order for prediction
- Make Design

Unique Trials for 3 Response-Surface Designs and # Quadratic Model Terms vs. # Continuous Factors



How many folks have any of these issues?

- Work with these different kinds of control variables/factors:
  - **Continuous/quantitative?** (Finely adjustable like *temperature*, *speed*, *force*)
  - **Categorical/qualitative?** (Comes in types, like material = *rubber*, *polycarbonate*, *steel* with mixed # of levels; 3 chemical agents, 4 decontaminants, 8 coupon materials…)
  - **Mixture/formulation?** (Blend different amounts of *ingredients* and the process performance is dependent on the *proportions* more than on the amounts)
  - **Blocking?** (e.g. "lots" of the same raw materials, multiple "same" machines, samples get processed in "groups" – like "eight in a tray," run tests over multiple days – i.e. variables for which there *shouldn't* be a causal effect
- Work with **combinations of these four kinds** of variables?
- Certain **combinations cannot be run**? (too costly, unsafe, breaks the process)
- Certain factors are **hard-to-change** (temperature takes a day to stabilize)
- Would like to **add onto existing trials**? (really expensive/time consuming to run)
- Characterize process or **run experiments using computer simulations**? (war gaming, agent-based, discrete event, computational fluid dynamics (CFD))
- Measure response **data in vicinity of physical limits**? (counts, hardness, resistivity can't fall below zero, or percentage yield or killed can't exceed 100%)

Create a Complex DOE

- Example of a complex design combining 4 types of factors with additional constraints
  - PDF of detailed steps available for you to follow
- Enter Factor Information
  - Factors and ranges (or levels): See next slide
- Propose Model
  - 2^{nd} order for prediction
- Make Design

Create a Sequential Custom DOE

- Enter Factor Information
  - Factors and ranges (or levels): See next slide
- Propose Models
  - 1st order for screening
  - 2nd order for prediction
- Make Designs – augment first design to support more complex model

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

Six factors described on page 9 of "Efficient M&S…" White Paper*

Variable	# Levels	Levels
Agent Codes	6	A, N, T, H, R, Y (categorical)
Season	3	Winter, Summer, Spring/Fall (categorical)
Time of Attack	3	0500, 1200, 2200 Local Time (continuous/categorical)
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/categorical)
Height of Burst	2	0, 10 m (continuous/categorical)
Total Cases	648	

*Currently fourth on list with date of June 2009
http://www.jmp.com/software/whitepapers/

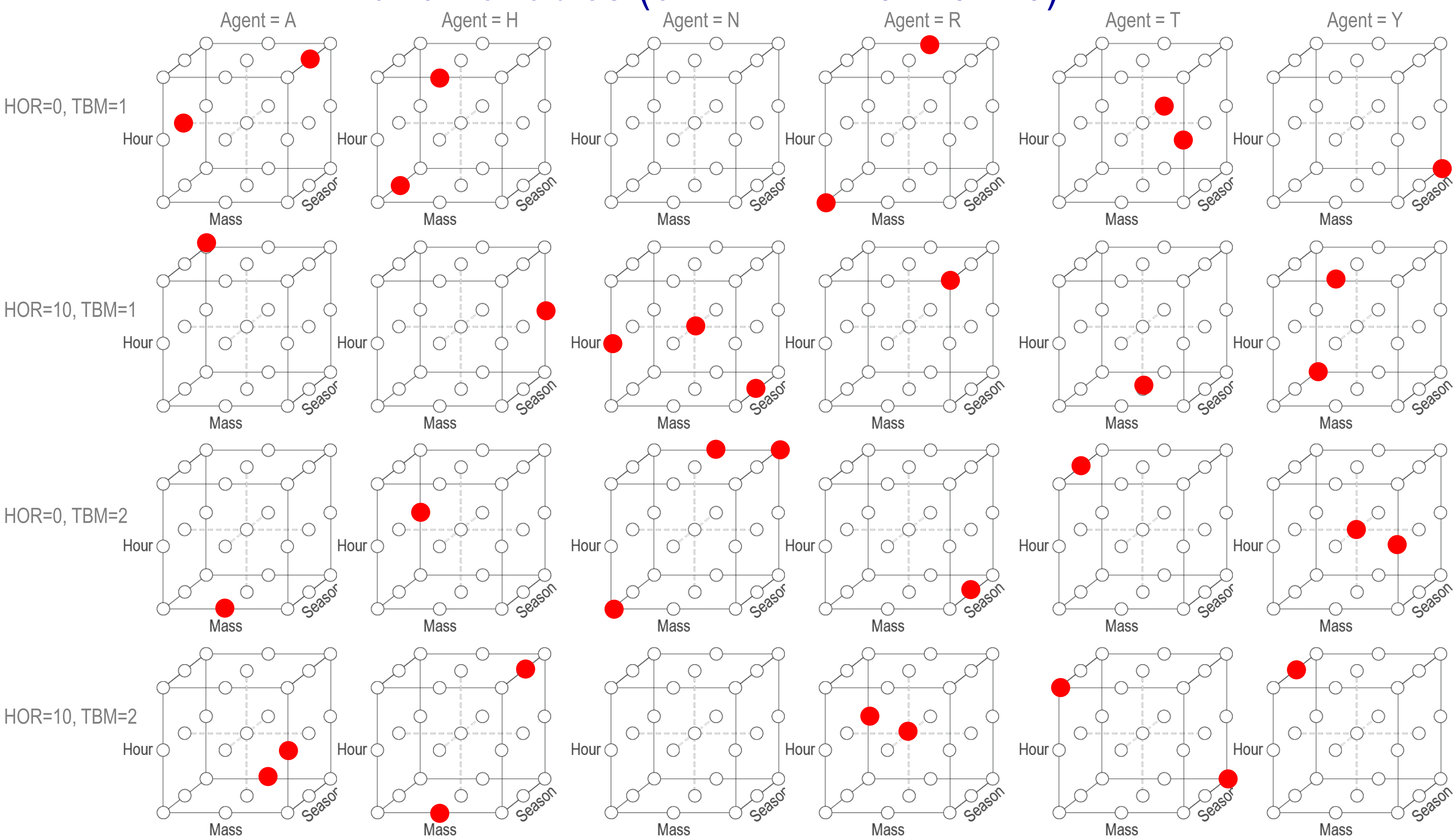
Four Stage Design Sequence

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

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

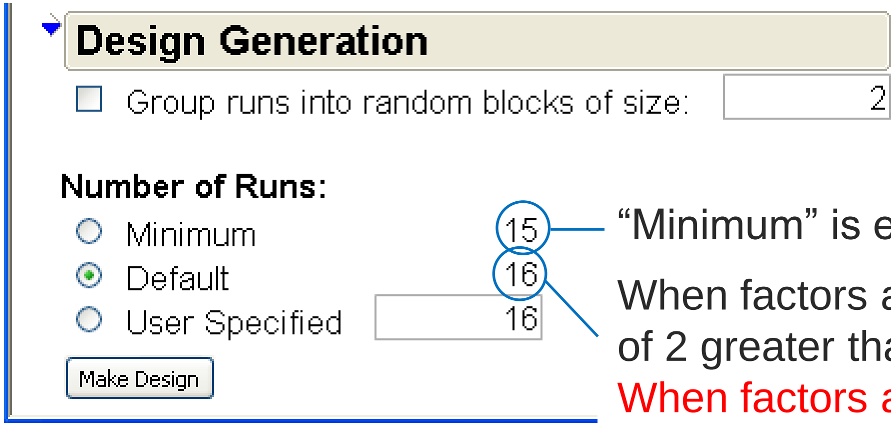
NOTE: Length of this green box should be longer than shown

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 of a Custom Design Analyzed for Stage 1

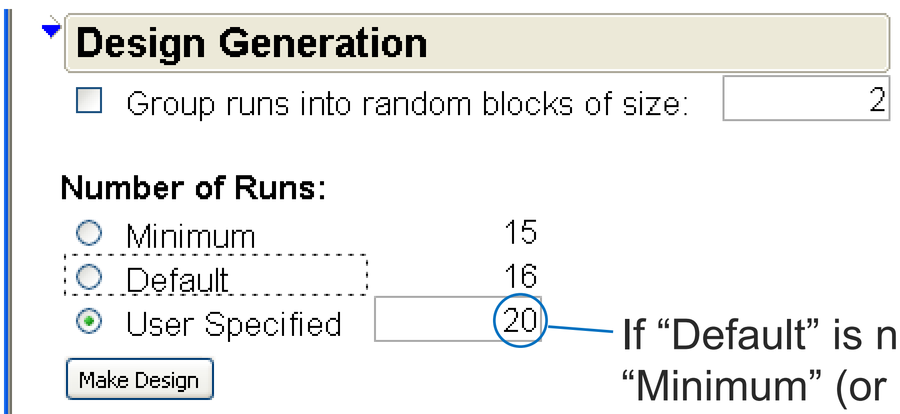
Increase degrees of freedom in model error estimate



"Minimum" is equal to number of terms in the model

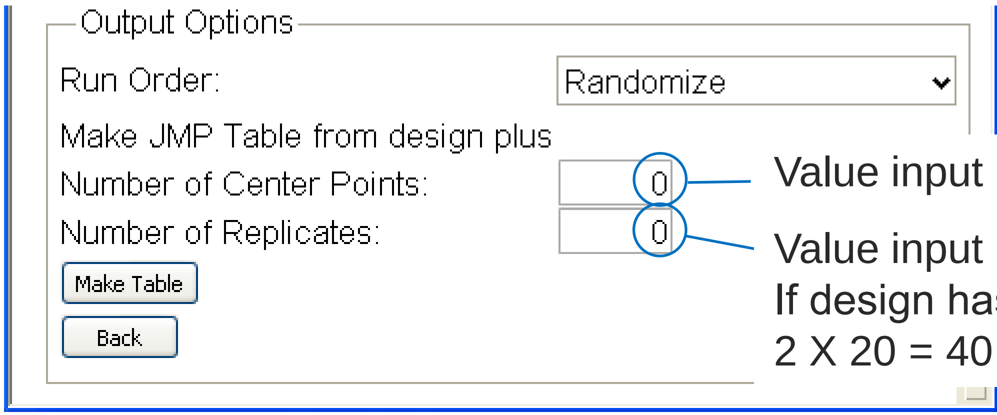
When factors are all *continuous* "Default" is the smallest power of 2 greater than the number of terms in the model

When factors are *categorical* "Default" is the smallest number evenly divisible by all numbers of factor levels (> minimum)



If "Default" is not at least 5 more than "Minimum," then enter 5 + "Minimum" (or more if you can afford it) in "User Specified"

Increase degrees of freedom for pure error estimate



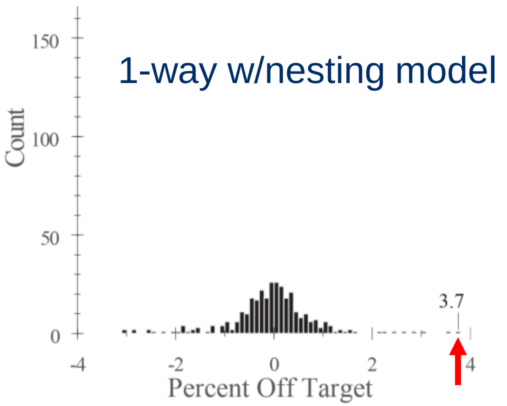
Value input is actual number of trials added to design.

Value input is number of times the design is replicated. If design has 20 unique trials, then a “2” here adds 2 X 20 = 40 more trials to design for a total of 60.

Having a model error estimate and a pure error estimate allows for a lack-of-fit test to be conducted.

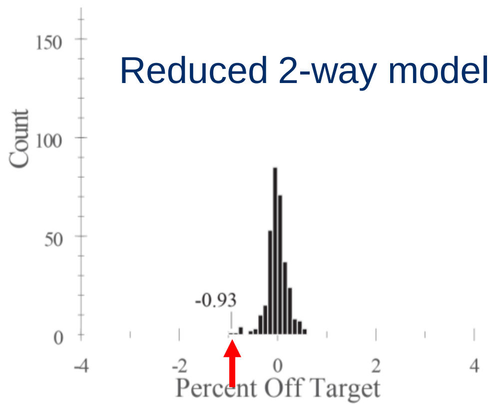
"Factor Sparsity" and "Effect Heredity" Used to Enhance Model Complexity

36 trials



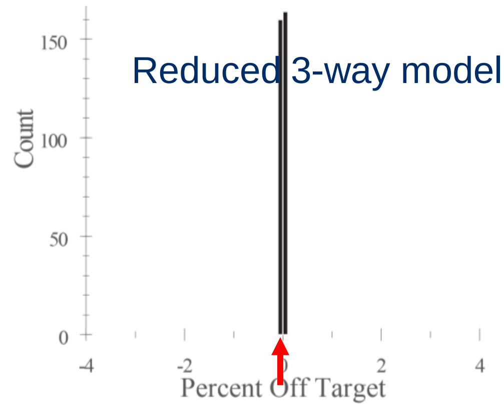
Worst Case = 3.7%
Half of Cases < 0.37%

108 trials

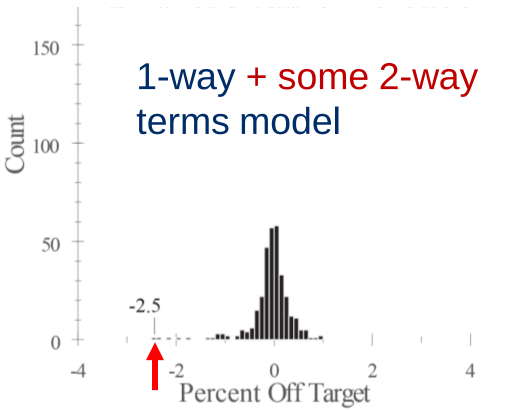


Worst Case = -0.93%
Half of Cases < 0.11%

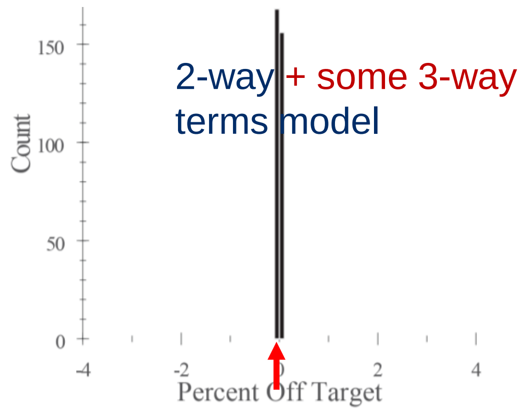
324 trials



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

Discussion Points on Integrated Test

- CT > DT > OT “use all the data” – IPTs
- TEWIPT – “If you can’t give me a requirements doc., I can’t & won’t design & cost the deliverable”
- Sooner a process is more fully characterized, the better – the earlier changes are made, the bigger the savings – use DOE as early as possible
- Full product may take years – need to characterize components to meet requirements
- Combine M&S with real testing – see papers by Wu, Higdon. e.g. breeze tunnel & CFD

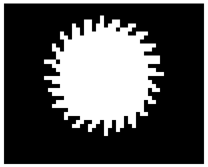
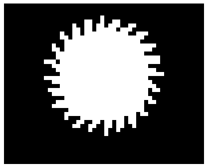
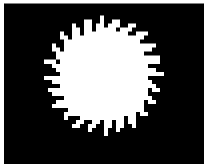
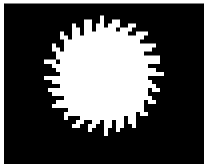
Army Conference on Applied Statistics

- Putting this year's focus on DOE & integrated T&E
- Probable management track – not the nuts and bolts but the what can it do and how do I know if someone is faking it (see Greg Hutto)
- Please pick up the flyer

Key Take Away Points

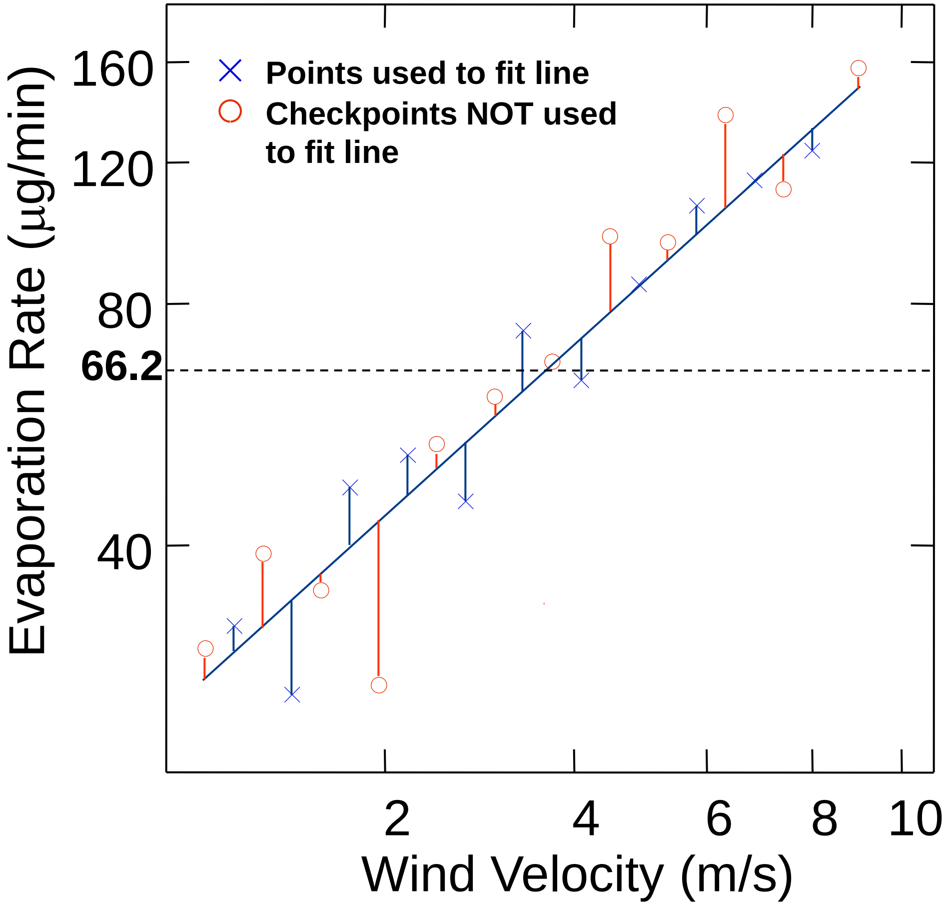
- Building predictive models of multiple responses provides decision makers with the knowledge to make better informed judgments and trade-offs.
- A Design of Experiments (DOE) is a collection of trials built to support a proposed model.
- An algorithmic design tool can quickly build a DOE for your predictive model
  - subject to real-world physical constraints
  - sequence of DOEs can be built to integrate testing
- Tools are great, but education is more important!
  "Successful use of DOE will require a cadre of personnel… with the professional knowledge and expertise…"

Backup Slides

MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
				
Sunny Hi: 42 F Lo: 25 F	Sunny Hi: 42 F Lo: 33 F	Sunny Hi: 49 F Lo: 33 F	Showers and mild Hi: 52 F Lo: 30 F	Sunny and pleasant Hi: 57 F Lo: 39 F

- A design run over 5 days that is sensitive to humidity might SHIFT on Thursday
  - What if because of the rain the tester from days 1-3 couldn't make it to work?
  - Or what if on that day there was a special visitor to the lab, or the power went out, or…?
- The block variable doesn't tell you the cause of the effect, just that a shift has been detected among blocks.
- The only way to be sure that no "unknown" factor has crept into the experiment is to test for it and "blocking" your design is not expensive.
- A block variable is a qualitative factor that has only 1-way effects (no interactions)
- If block variable shows no effect it can be deleted from the analysis.

Comparing Residual SD, *Checkpoint RMS* and Raw SD for a Single Control Variable

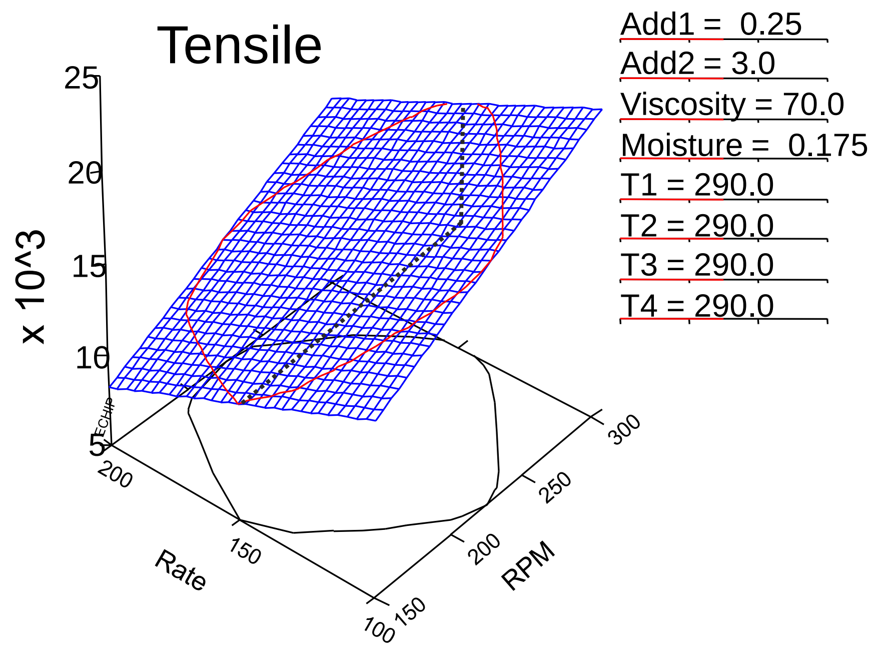



```
N trials          = 11
N terms           = 2
Residual DF       = 9
Residual SD       = 0.0655
(Model error)

N checkpoints     = 12
Checkpoint RMS    = 0.0808
(Prediction error)

Raw SD            = 0.2245
(Error about mean of data
 for 11 trials used to fit)
```

Graph paper used has log10 vertical scale and cube-root horizontal scale.
(NOTE: Real Data NOT Used)

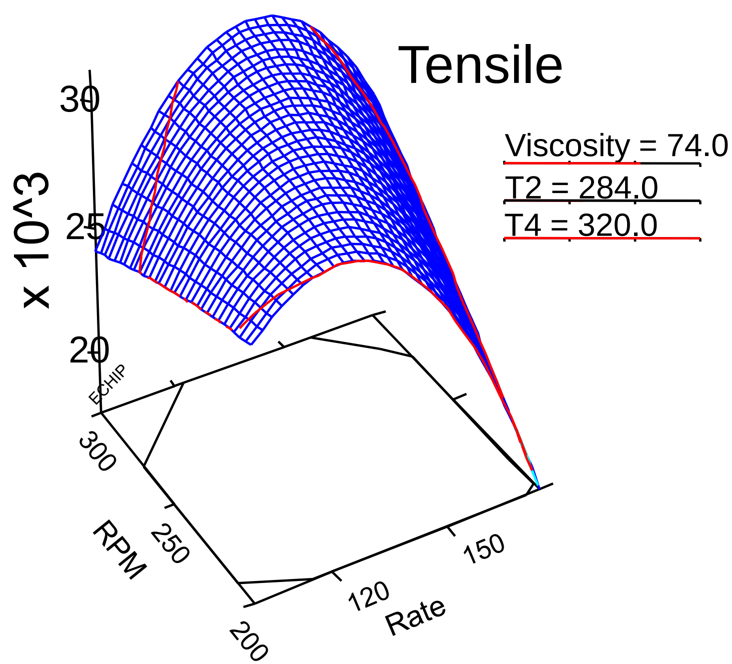
Screening 10-var.

Can't draw 10-D cube with $2^{10} = 1024$ corners! Design only needs ≈ 2% ≈ 20 trials*



Response-Surface 5-var.

Can't draw 5-D cube with $3^5 = 243$ candidate trials! Design has 27 trials ≈ 11%



*PB12, FF16, MM17, PB20, PB24FO, FF32