MIAMI UNIVERSITY

VCU

A Meta Analysis of Response Surface Designs

Presented by: **Maria Weese**

Rebecca Ockuly

Byran Smucker

David Edwards

Le Chang

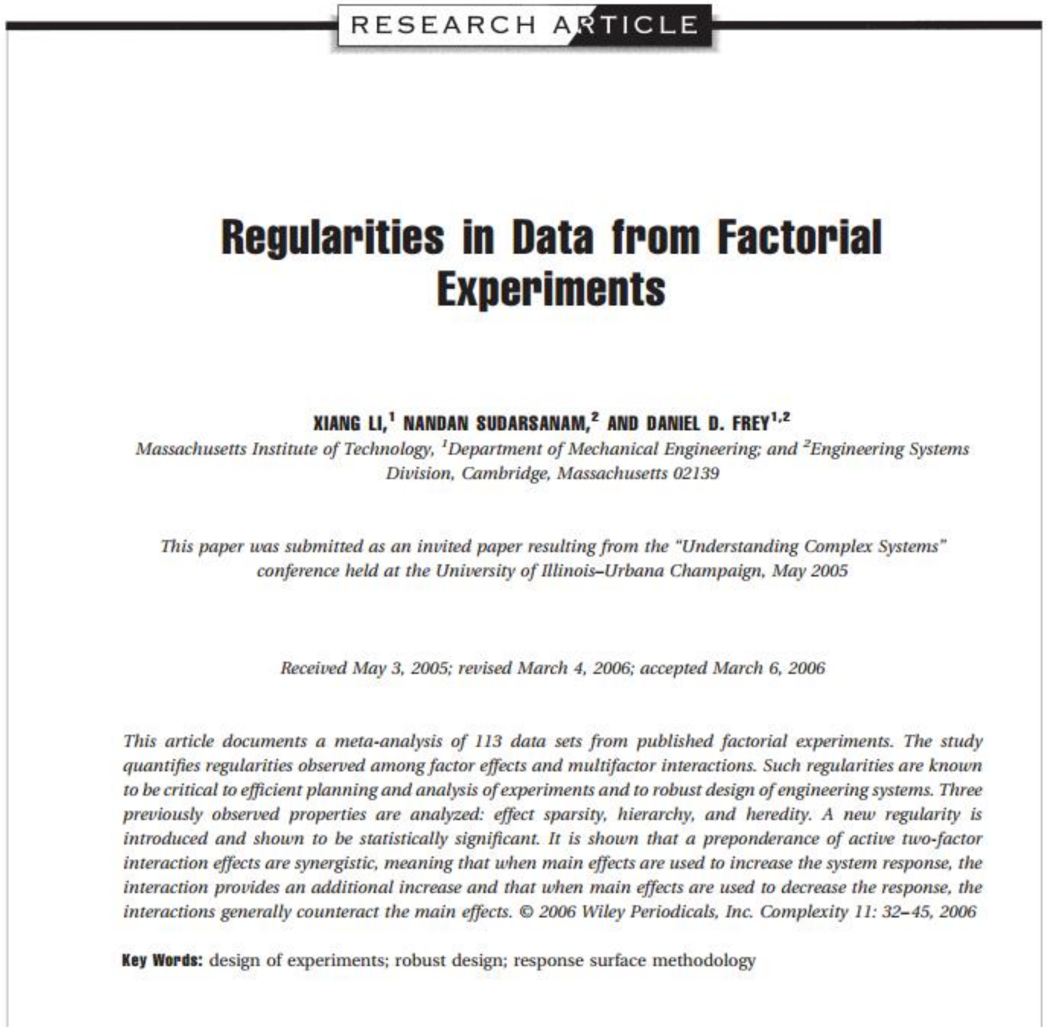
FARMER SCHOOL OF BUSINESS
Information Systems & Analytics


THE TRUTH IS OUT THERE

Motivation

Li, Sudarsanam and Frey (2006) give great insight regarding effect sparsity, heredity, and hierarchy for factorial designs.

They also provide information regarding the signs of the interaction terms.



RESEARCH ARTICLE

Regularities in Data from Factorial Experiments

XIANG LI,[1] NANDAN SUDARSANAM,[2] AND DANIEL D. FREY[1,2]
Massachusetts Institute of Technology, [1]Department of Mechanical Engineering; and [2]Engineering Systems Division, Cambridge, Massachusetts 02139

This paper was submitted as an invited paper resulting from the "Understanding Complex Systems" conference held at the University of Illinois–Urbana Champaign, May 2005

Received May 3, 2005; revised March 4, 2006; accepted March 6, 2006

This article documents a meta-analysis of 113 data sets from published factorial experiments. The study quantifies regularities observed among factor effects and multifactor interactions. Such regularities are known to be critical to efficient planning and analysis of experiments and to robust design of engineering systems. Three previously observed properties are analyzed: effect sparsity, hierarchy, and heredity. A new regularity is introduced and shown to be statistically significant. It is shown that a preponderance of active two-factor interaction effects are synergistic, meaning that when main effects are used to increase the system response, the interaction provides an additional increase and that when main effects are used to decrease the response, the interactions generally counteract the main effects. © 2006 Wiley Periodicals, Inc. Complexity 11: 32–45, 2006

Key Words: design of experiments; robust design; response surface methodology

Motivation

However, we were not able to answer the following questions from their study:

1. How often are second order effects found to be active?
2. Is there a relationship between the active second order effects and active main effects and active interactions?
3. How large are second order effects relative to main effects? Interactions?
4. Does effect sparsity differ with second order designs?

The answers to these questions will inform us on how to design more realistic simulation scenarios, give information about favorable aliasing structure, and information as to what experimenters can expect to find from standard second order designs.

Terms

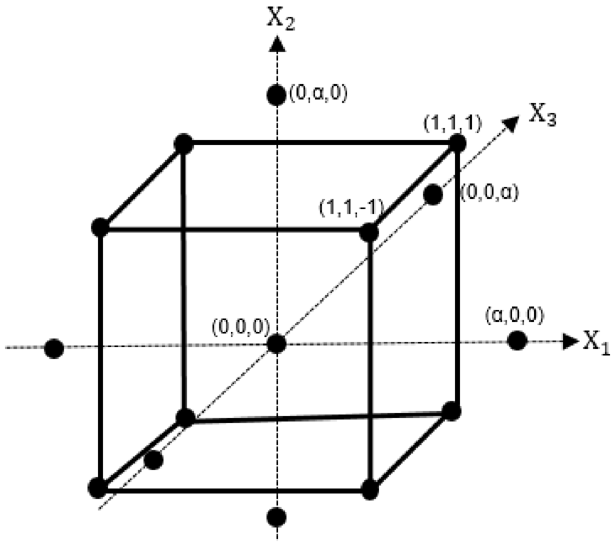
Effect Sparsity: Only a small proportion of the effects in an experiment will actually have a significant impact on the response.

Effect Hierarchy: Lower-order effects are more likely to be important than higher-order effects, hence, the magnitude of lower order effects will generally be increased.

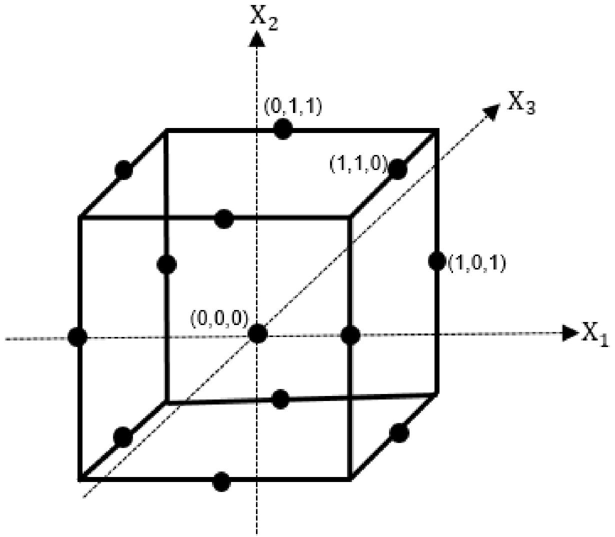
Effect Heredity: In order for an interaction to be active, at least one of its “parent” factors should also be active.

Quantifying Second Order Effects

Response surface designs are able estimate second order effects. Most commonly experimenters use either a Central Composite Design (CCD) or a Box Behnken Design (BBD).



CCD



BBD

Wu, CFJ, and Michael Hamada. 2009. *Experiments: Planning, Analysis, and Optimization. Wiley Series in Probability and Statistics*. Wiley.

Obtaining the “Population”

We used the Web of Science API to search all journal articles with the following characteristics:

- Published between January 1st 1990 and December 31st 2014 in the Science Citation Index and the Social Sciences Citation Index.
- Searched on terms “Response Surface” OR “Central Composite” OR “Box Behnken” OR “Box-Behnken”.

This returned 24,286 search results from which we extracted the citations. We used a stratified random sample from this population.

The Sample

134 Papers:

Design type: 106 CCD's and 28 BBDs

Screening: 48 mentioned a screening experiment, 86 did not

Coding: 80 used coded units, 54 uncoded units

Number of Factors	Frequency (by Paper)	Frequency (by Response)
2	16	48
3	43	94
4	28	40
5	39	71
6	6	7
7	2	3
Total	134	263

The Sample

263 Responses:

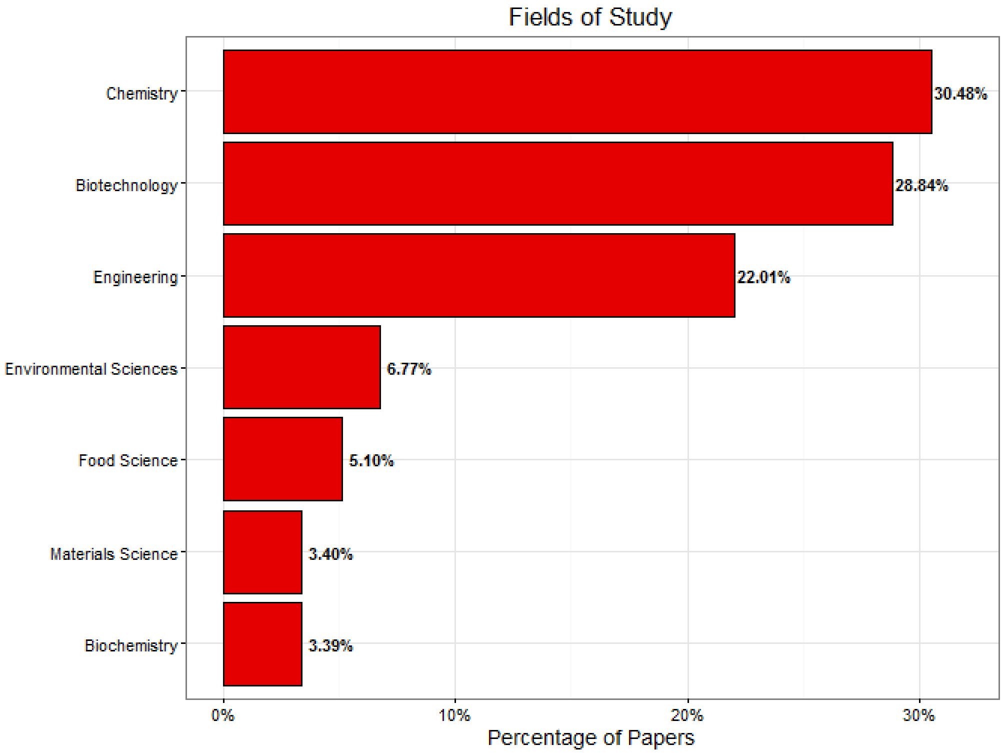
Runs: 9 to 100 with a mode=6 and a median=20

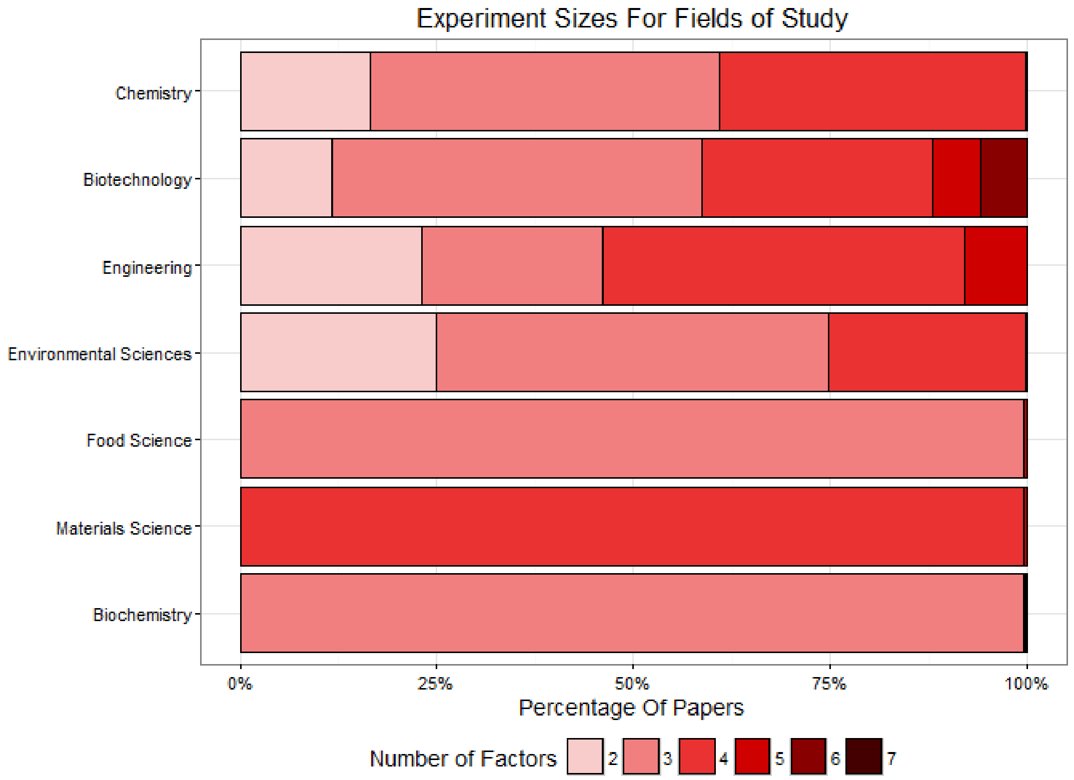
Axial distance (for CCDs): 0.6 to 2.83, mode=2, median=1.69

Lack of fit: 89 show LOF, 166 do not, and LOF could not be estimated in 8 models.

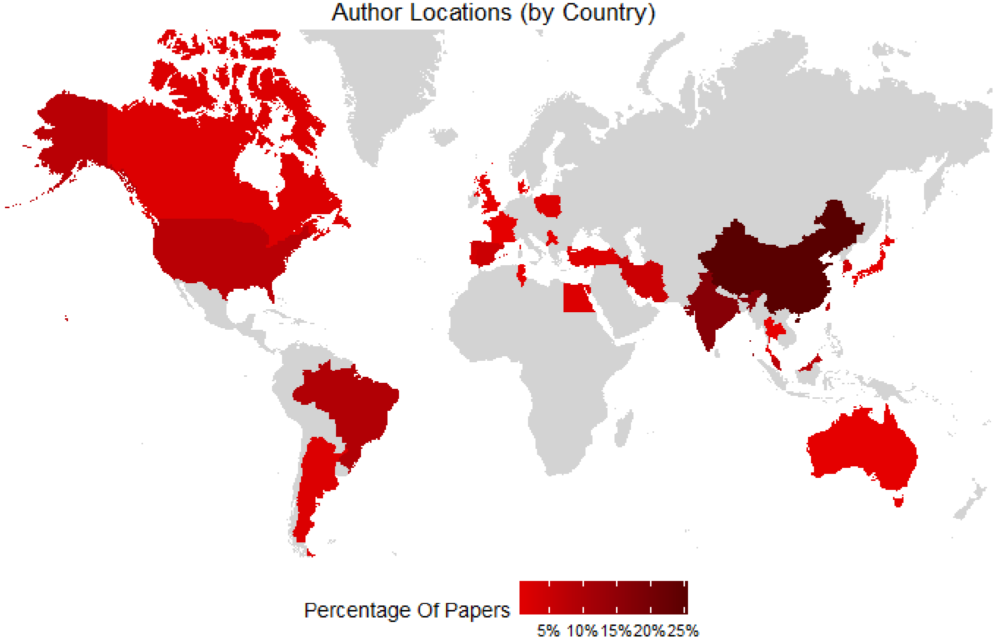
Number of Factors	Frequency (by Paper)	Frequency (by Response)
2	16	48
3	43	94
4	28	40
5	39	71
6	6	7
7	2	3
Total	134	263

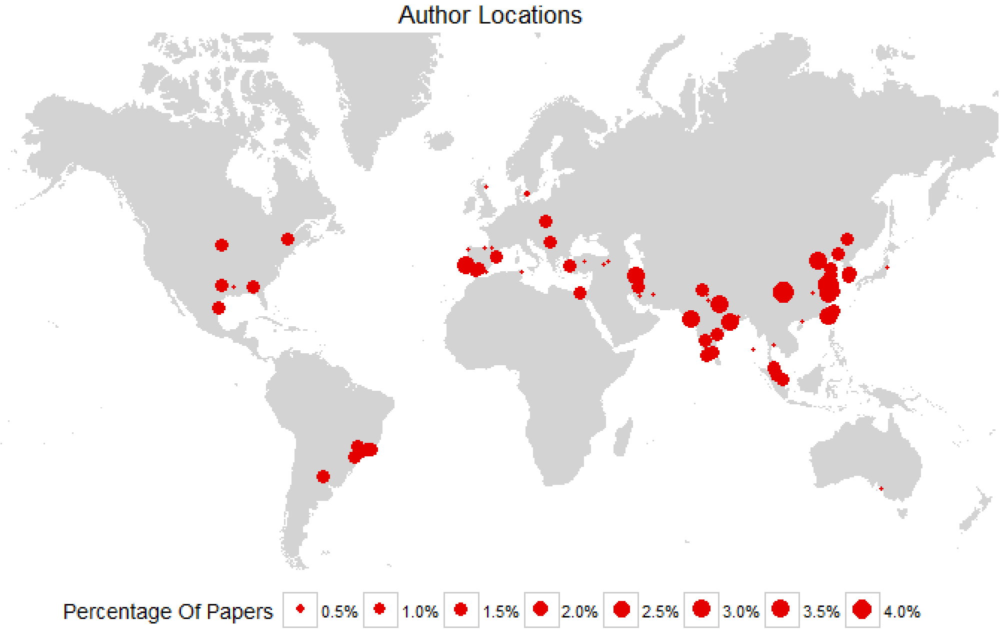
Exploratory Data Analysis





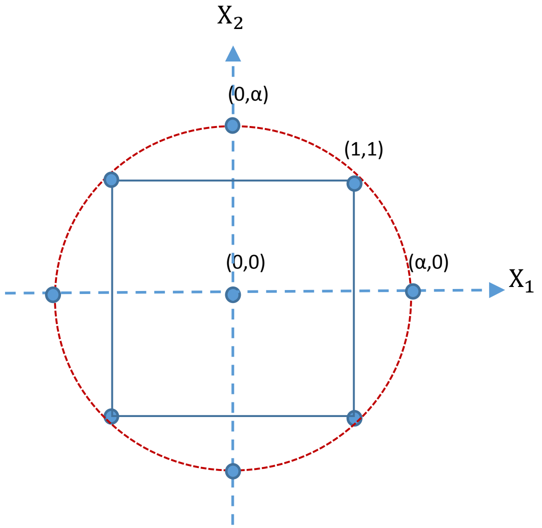
Exploratory Data Analysis



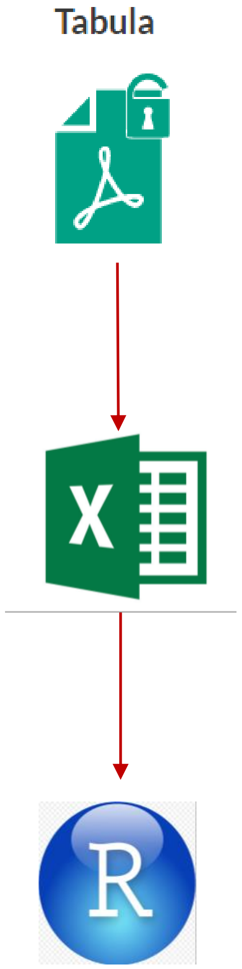


Example

CCD with 2 factors and 1 response and n= 13 runs.



	Factors		Response
	NaCl (X1)	**CaSO (X2)**	**Lipase Activity (Y)**
Factorial Points	-1	-1	2.8
	-1	1	2.6
	1	-1	2.9
	1	1	2.2
Axial Points	-1.414	0	2.1
	1.414	0	2.4
	0	-1.414	2.7
	0	1.414	2
Center Points	0	0	3.3
	0	0	2.9
	0	0	3.2
	0	0	3.5
	0	0	3.3



Liu, Chien-Hung, Wei-Bin Lu, and Jo-Shu Chang. 2006. "Optimizing Lipase Production of *Burkholderia* Sp. by Response Surface Methodology." *Process Biochemistry* 41 (9). Elsevier: 1940–44.

Example, cont.

We defined an active effect in two ways: Using a *p-value*<0.05 and using a False Discovery Rate (FDR) adjusted *p-value*<0.05. We used the set of identified active effects to quantify effect heredity.

We used the absolute value of the t-statistics to quantify effect hierarchy and the signs of the active interactions and main effects to establish interference or reinforcement.

We recorded the p-value for the LOF test and the stationary point.

	Estimate	Std. Error	t-value	p-value	FDR p-value
Intercept	3.240	0.116	27.955	1.925e-08	1.155e-07
X1	0.016	0.092	0.170	0.870	0.870
X2	-0.236	0.092	-2.578	0.037	0.056
X1:X2	-0.125	0.130	-0.965	0.367	0.440
X1^2	-0.414	0.098	-4.210	0.004	0.012
X2^2	-0.364	0.098	-3.701	0.008	0.016

Multiple R-squared: 0.835, Adjusted R-squared: 0.717

	DF	Sum Sq	Mean Sq	F-value	p-value
FO(X1, X2)	2	0.448	0.224	3.338	0.096
TWI(X1, X2)	1	0.063	0.063	0.931	0.367
PQ(X1, X2)	2	1.870	0.935	13.919	0.004
Residuals	7	0.470	0.067		
Lack of Fit	3	0.278	0.093	1.932	0.266
Pure Error	4	0.192	0.048		

Stationary Point: (0.070, -0.337)

Effect Sparsity

p-value < 0.05

Effect Type	Proportion Active	Standard Error
Main Effects	0.58	0.03
Interaction Effects	0.21	0.03
Quadratic Effects	0.45	0.03

FDR p-value < 0.05

Effect Type	Proportion Active	Standard Error
Main Effects	0.52	0.03
Interaction Effects	0.16	0.02
Quadratic Effects	0.38	0.03

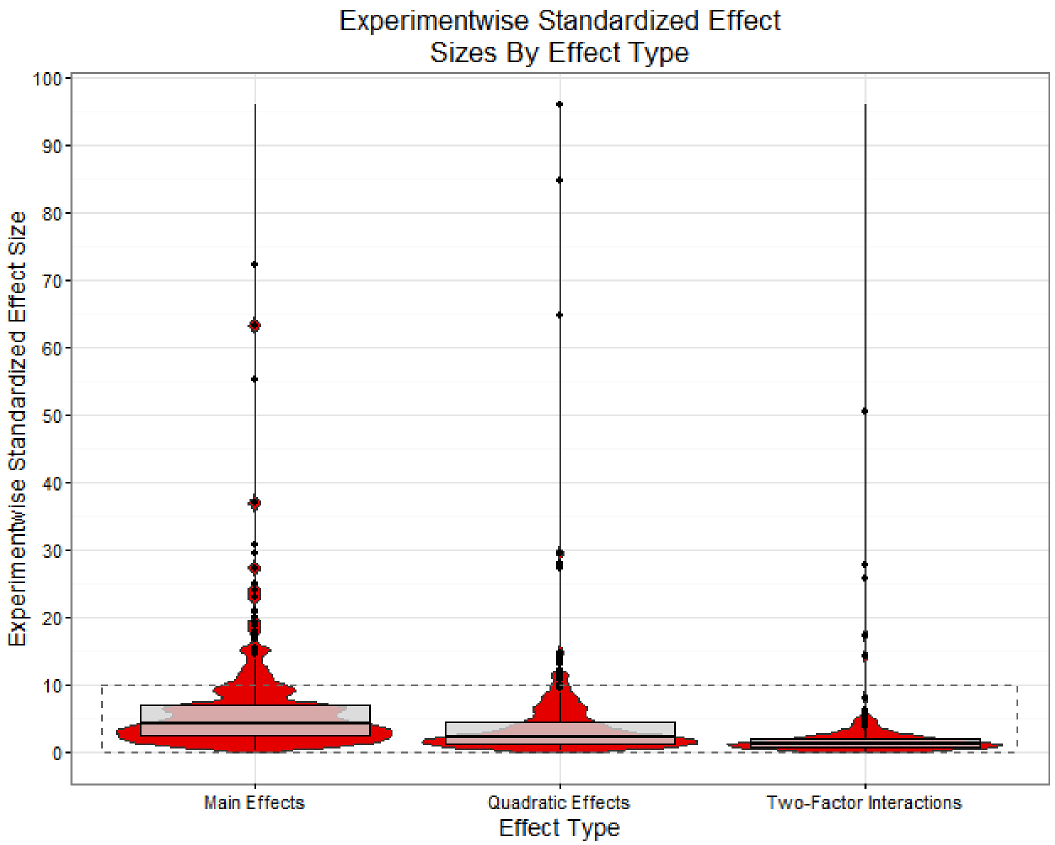
- These proportions are adjusted according to the stratified sampling.
- More quadratic than interaction effects are active.
- Li, Sudarsanam and Frey (2006) found 41% of main effects to active and 11% of two factor interactions to be active.

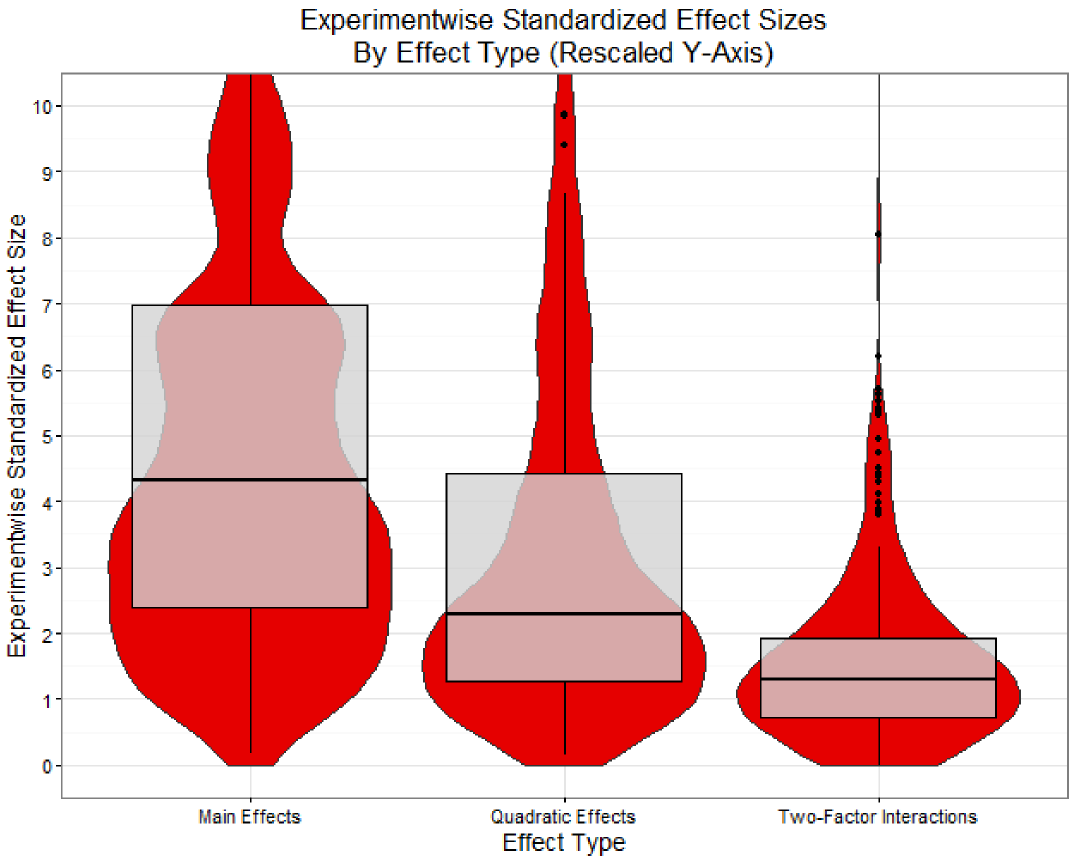
Effect Hierarchy

Effect Type	Mean Effect Size	Standard Error	Weighted Median
Main Effects	6.25	0.47	3.474
Interaction Effects	1.60	0.10	0.935
Quadratic Effects	3.59	0.21	1.956

- These averages reflect the stratified sampling procedure.
- All effects are included, not just active effects.
- The main effects are largest, however, the quadratic effects are still larger than the interaction effects.
- Li, Sudarsanam and Frey (2006) reported that the median of the main effects to be about 4 times larger than the median of the two factor interactions.

Effect Sparsity





Interaction Effect Heredity

$\hat{p}$(*involved main effects active* | *active Interaction*)

p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A, B	0.65	0.06
Weak	AB, A	0.30	0.06
None	AB	0.05	0.03

FDR p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A, B	0.69	0.06
Weak	AB, A	0.25	0.05
None	AB	0.06	0.03

- These proportions are adjusted to reflect the stratified sampling.
- The majority of active interactions have strong heredity, but weak heredity is not uncommon.
- Li, Sudarsanam and Frey (2006) reported

  $\hat{p}$ (*interaction is active* | *main effects are active*)=33% (strong), 4.5% (weak) and 0.48% (none)

Quadratic Effect Heredity

$\hat{p}(\textit{involved main effect active} \mid \textit{active quadratic effect})$

p-value < 0.05

Strength	Example	Proportion	Std. Error
Heredity	A^2, A	0.73	0.05
No Heredity	A^2	0.27	0.05

FDR p-value < 0.05

Strength	Example	Proportion	Std. Error
Heredity	A^2, A	0.72	0.05
No Heredity	A^2	0.28	0.05

- These proportions are adjusted to reflect the stratified sampling.
- Most of the active quadratic effects have their corresponding main effect also active.
- FDR adjusted p-values do not seem to make a difference in the proportions.

Quadratic Interaction Effect Heredity

$\hat{p}(\textit{number of quadratic effects active } | \textit{active interaction})$

p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A^2, B^2	0.46	0.06
Weak	AB, A^2	0.38	0.06
None	AB	0.16	0.04

FDR p-value < 0.05

Strength	Example	Proportion	Std. Error
Strong	AB, A^2, B^2	0.49	0.06
Weak	AB, A^2	0.33	0.06
None	AB	0.18	0.05

- These proportions are adjusted according to the stratified sampling.
- The majority of active interactions have strong quadratic heredity.
- An active interaction with none of the parent quadratic effects is not uncommon.

Expanded Effect Heredity

		p-value < 0.05		FDR p-value < 0.05	
Strength	**Example**	**Proportion**	**Std. Error**	**Proportion**	**Std. Error**
Strong-Strong	AB, A, B, A^2, B^2	0.31	0.06	0.33	0.06
Strong-Weak	AB, A, B, B^2	0.25	0.05	0.25	0.06
Strong-None	AB, A, B	0.09	0.03	0.12	0.04
Weak-Strong	AB, B, A^2, B^2	0.14	0.04	0.14	0.04
Weak-Weak (Same)	AB, A, A^2	0.10	0.04	0.05	0.03
Weak-Weak (Different)	AB, B, A^2	0.02	0.02	0.01	0.01
Weak-None	AB, A	0.04	0.02	0.04	0.03
None-Strong	AB, A^2, B^2	0.01	0.01	0.03	0.02
None-Weak	AB, A^2	0.01	0.01	0.01	0.01
None-None	AB	0.03	0.02	0.01	0.01

- 80% (p-value <0.05) and 87% (FDR p-value<0.05) of the active two-factor interactions exhibit at least one strong component of expanded heredity.
- These proportions are adjusted to reflect the stratified sampling.

Reinforcement and Interference

We examined the signs of the interaction effects and the associated interactions and classified them according to the following definitions (Kunter et al 2005):

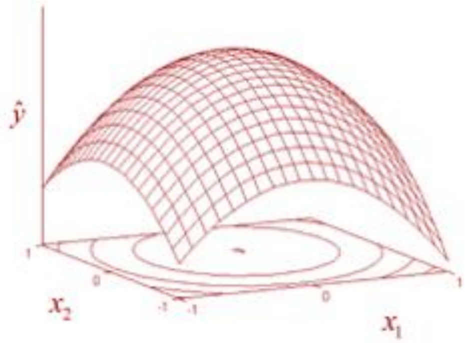
Reinforcement: Both parent factors associated with the two-factor interaction have the same sign as the interaction effect. **28.11%**

Interference: Both parent factors associated with two-factor interactions have opposite sign from the interaction effect. **27.76%**

Other: One parent factor has a positive effect and the other factor is negative (or visa versa). **44.13%**

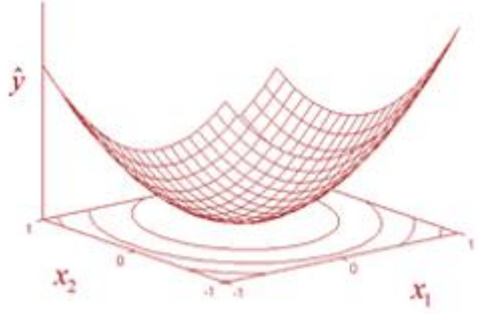
Note, this not the same measure Li, Sudarsanam and Frey (2006) quantified.

Stationary Points



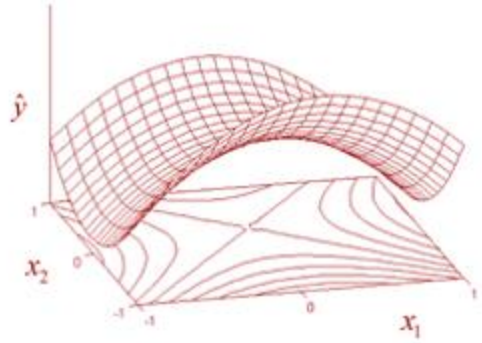
Maximum: All eigenvalues negative in canonical analysis

23 experiments



Minimum: All eigenvalues positive in canonical analysis

43 experiments



Saddle Point: Mixture of positive and negative eigenvalues

197 experiments

http://reliawiki.org/index.php/Response_Surface_Methods_for_Optimization

Caveats

By treating experiments with multiple responses as single experiments there is a potential to violate the assumption of independence.

Our "population" that we generalized to is only published experiments from Web of Science with all experimental data given in the paper. Weights were calculated accordingly.

The File Drawer Problem

Conclusions

Effect Sparsity

- Over half of all main effects were active (58%, 52%)
- Many quadratic effects were found active (45%, 38%)
- Interactions made up a much smaller proportion of the active effects (21%, 16%)

Effect Hierarchy

- Main effects were largest, followed by quadratic then interaction effects.

Conclusions

Effect Heredity

- Most interactions exhibit strong heredity, 2:1 strong to weak.
- Most active quadratic effects have the main effect also active (73%, 72%)
- Almost half of all active interactions have strong heredity with both quadratic and main effects.
- Strong heredity in general (both quad and interaction) dominate fitted models.

Interaction Signs

- There is a fairly equal division between reinforcing and interfering interactions.

Comments or Suggestions?

For a copy please email: weeseml@miamioh.edu

References

Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1). Royal Statistical Society, Wiley: 289–300. http://www.jstor.org/stable/2346101.

Kutner, M.H. 2005. Applied Linear Statistical Models. McGraw-Hill/Irwin Series Operations and Decision Sciences. McGraw-Hill Irwin.

Lenth, Russell V. 2009. “Response-Surface Methods in R, Using rsm.” *Journal of Statistical Software* 32 (7): 1–17. http://www.jstatsoft.org/v32/i07/.

Li, Xiang, Nandan Sudarsanam, and Daniel D Frey. 2006. “Regularities in Data from Factorial Experiments.” *Complexity* 11 (5): 32–45.

Liu, Chien-Hung, Wei-Bin Lu, and Jo-Shu Chang. 2006. “Optimizing Lipase Production of *Burkholderia* Sp. by Response Surface Methodology.” *Process Biochemistry* 41 (9). Elsevier: 1940–44.

Lohr, S.L. 1999. *Sampling: Design and Analysis*. Sampling: Design and Analysis, v. 1. DuxburyPress.

Wu, CFJ, and Michael Hamada. 2009. *Experiments: Planning, Analysis, and Optimization. Wiley Series in Probability and Statistics*. Wiley.