Topics in Advanced Experimental Design

Nonclinical Biostatistics Conference

Steve Buyske

Rutgers University

June 12, 2017

Classical Response Surface Methodology

Optimal Design

Definitive Screening Designs

Experimental Design for Mixtures

Analysis of Mixture Data Using Partial Least Squares

Chemometrics-style “Multivariate Design”

Multiple Responses and the Design Space

Bayesian predictive examples

Classical Response Surface Methodology

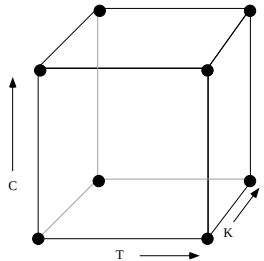
A set of designs and methodologies to optimize a process or mixture, with steps to

- Identify the primary factors
- Identify the region of interest for the primary factors
- Develop *useful* models in the region of the optimum settings of the factors
- Confirm the results

Classical Response Surface Methodology: Fractional Factorials

Two-level factorial designs, 2^k, k the number of factors, are the basic building blocks

- can reveal interactions of factors
- provide excellent precision for estimates
- give some sense of range of validity for conclusions about each factor
- do not allow for estimation of quadratic effects
- can be modified to reduce the number of the runs as well as increase the size of the design space
- are excellent building blocks



Classical Response Surface Methodology: Fractional Factorials

Fractional factorials are often run without replications
To pick *important* terms,

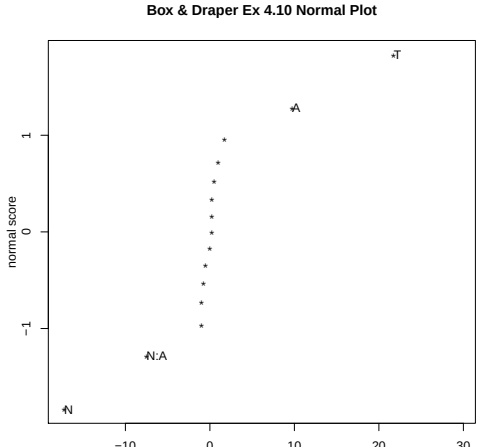
- Normal plot
- Half-normal
- Lenth test

Example (from Box & Draper, Prob 4.10): 4 factors (catalyst concentration, NaOH, level of agitation, temperature) in unreplicated 2^4 factorial. The response variable is the yield of an unspecified chemical process

Classical Response Surface Methodology: Fractional Factorials

Normal Plot

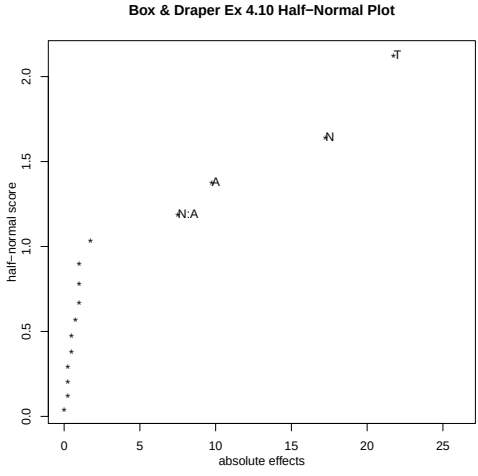
Idea: - Most of the estimated effects will just normal noise—all with the same variance
- Important effects will look like outliers



Classical Response Surface Methodology: Fractional Factorials

Half-Normal Plot

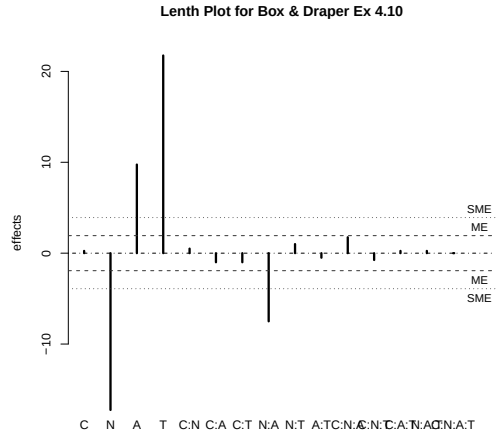
Same-idea, but take advantage that sign shouldn't be relevant



Classical Response Surface Methodology: Fractional Factorials

Lenth Test

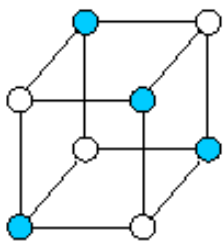
Calculate a pseudo standard error and compare ratios to reference distribution



Classical Response Surface Methodology: Fractional Factorials

Blocking in Fractional Factorials

- May need to run smaller sets of runs than will fit in full factorial
- Would like any block effect to be orthogonal to effect estimates
- Basic idea: define blocks by signs of high-order interaction
  - E.g., 2^3 with two blocks, define block by sign of $A \times B \times C$ (using ± 1 coding)
  - Block will be confounded with the ABC interaction
  - E.g., 2^3 with four blocks, define blocks by the pair of signs of $A \times B$ and $A \times C$
  - Block will be confounded with the AB, AC, and BC interactions



Classical Response Surface Methodology: Fractional Factorials

Fractional Factorial Designs

When K becomes large, the number of experimental runs can become unmanageable.

- E.g., $k = 7$, $2^k = 2^7 = 128$
- What do we get for all this work? Estimates of all main effects and interactions. With $K = 7$

Interactions	number	
grand mean	1	$\binom{7}{0}$
main effects	7	$\binom{7}{1}$
2-way	21	$\binom{7}{2}$
3-way	35	$\binom{7}{3}$
4-way	35	$\binom{7}{4}$
5-way	21	$\binom{7}{5}$
6-way	7	$\binom{7}{6}$
7-way	1	$\binom{7}{7}$

- Most higher order interactions are not likely to be real
- A lot of effort for an iffy part of model

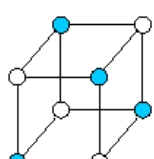
Classical Response Surface Methodology: Fractional Factorials

Imagine that we our we had run our 2^7 in 2 blocks, defined by the sign of

$$ABCDEFG$$

,

- Now imagine running just the block $ABCDEFG = 1$
- We have a design equivalent to a full factorial in 6 factors, with the levels of factor G determined by $G = ABCDEF$
  - G is confounded with $ABCDE$
  - all main effects confounded with 5 way interactions
  - all 2 way interactions confounded with 4 way interactions and so on
- Denote as a 2^{7-1} fractional factorial



Classical Response Surface Methodology: Fractional Factorials

Just as we can create $2, 4, 8, \ldots$ blocks, we can create half, quarter, eighth, ... fractions
For example,

- Define the levels of F and G by $F = ABCD$ and $G = BCDE$
- Gives a 2^{7-2} design
- There are implicit defining relations:
  - $I = ABCDF$ and $I = BCDEG \Rightarrow I = AEFG$
- The defining relations determine all of the confounding patterns—which effects are aliased with which
  - E.g., since $I = AEFG$, AF is completely confounded with EG
- We define resolution of a design is the length of the shortest word in a defining relation
  - E.g., the design here has resolution IV, written 2^{7-2}_{IV}

Classical Response Surface Methodology: Fractional Factorials

Continuing the example, we can construct a 2_{IV}^{7-3} design with these generators

- $E = ABC$
- $F = BCD$
- $G = ACD$

	A	B	C	D	E	F	G
1	-1	-1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	1	-1	1
3	-1	1	-1	-1	1	1	-1
4	1	1	-1	-1	-1	1	1
5	-1	-1	1	-1	1	1	1
6	1	-1	1	-1	-1	1	-1
7	-1	1	1	-1	-1	-1	1
8	1	1	1	-1	1	-1	-1
9	-1	-1	-1	1	-1	1	1
10	1	-1	-1	1	1	1	-1
11	-1	1	-1	1	1	-1	1
12	1	1	-1	1	-1	-1	-1
13	-1	-1	1	1	1	-1	-1
14	1	-1	1	1	-1	-1	1
15	-1	1	1	1	-1	1	-1
16	1	1	1	1	1	1	1

Classical Response Surface Methodology: Fractional Factorials

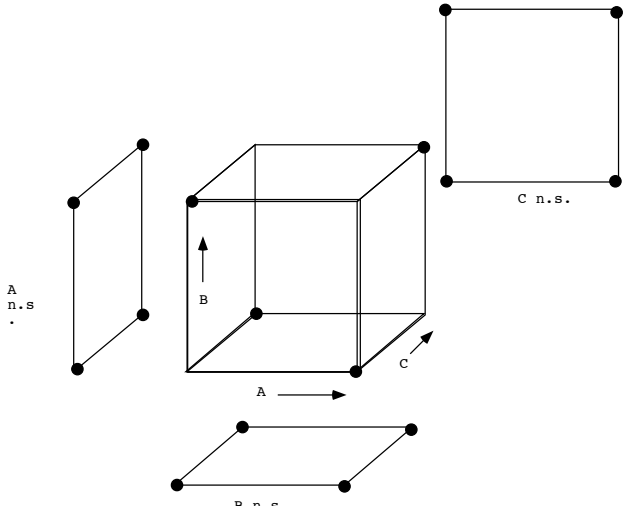
The generators $E = ABC$, $F = BCD$, $G = ACD$ then determine the defining relations

- $I = ABCE$
- $I = BCDF$
- $I = ACDG$
- $I = ADEF$
- $I = BDEG$
- $I = ABFG$
- $I = CEFG$

For a p-th fraction, there are 2^p defining relations (including $I = I$)

Classical Response Surface Methodology: Fractional Factorials

In a fractional factorial design, if any factor is inert the design projects to a fractional factorial in the remaining factors



Classical Response Surface Methodology: Fractional Factorials

Since

- a fractional factorial design is like a blocked full factorial with all but one block missing
- if the results are ambiguous we can *augment* the design with another of the blocks
  - typically be flipping the signs of some or all factors ("fold over")
- or a fraction of a missing block
- (or use optimal design)

Classical Response Surface Methodology: Fractional Factorials

Fold over example:

- Original design a 2^{5-2}_{III} with $D = AB$ and $E = BC$ and defining relations
  - $I = ABD = BCE = ACDE$
- We flip the signs (levels) of all the factors, so $D = -AB$ and $E = -BC$ so the the defining relations for the new block are
  - $I = -ABD = -BCE = ACDE$
- Since $I = ACDE$ is common to both halves the combined fractions have this as the relation
- Block will be confounded with 3 way interactions
- The original design plus the fold over form a blocked 2^{5-1}_{IV}

Classical Response Surface Methodology: Plackett-Burman Designs

Plackett and Burman are a popular alternative to fractional factorial designs for detecting main effects.

- Allow for the number of design points to be a multiple of 4
- Can have up (number of factors - 1) runs, or drop off extra columns from the design and gain some df
- No power of 2 restriction
- Aliasing pattern for a Plackett-Burman design is much more complex. Generally, each main effect is aliased with every 2-way interaction not involving that effect

Classical Response Surface Methodology: Plackett-Burman Designs

	A	B	C	D	E	F	G	H	I	J	K
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	-1	-1	-1	1	-1	-1	-1
3	1	1	-1	-1	-1	1	-1	-1	1	-1	1
4	1	-1	1	-1	1	1	1	-1	-1	-1	-1
5	1	-1	-1	1	-1	-1	1	-1	1	1	-1
6	1	-1	-1	-1	1	-1	-1	1	-1	1	1
7	-1	1	1	-1	-1	-1	1	-1	-1	1	1
8	-1	1	-1	1	1	1	-1	-1	-1	1	-1
9	-1	1	-1	-1	1	-1	1	1	1	-1	-1
10	-1	-1	1	1	1	-1	-1	-1	1	-1	1
11	-1	-1	1	-1	-1	1	-1	1	1	1	-1
12	-1	-1	-1	1	-1	1	1	1	-1	-1	1

A 12-run Plackett-Burman design

Classical Response Surface Methodology: Plackett-Burman Designs

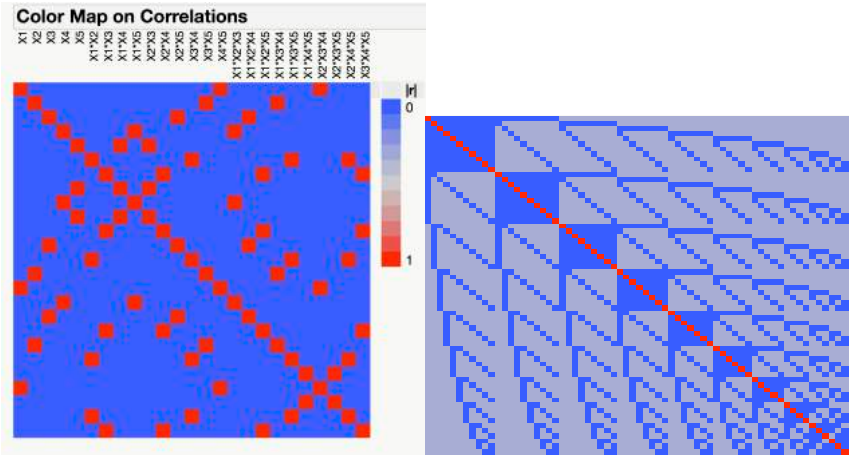
Reminder:

- From regression methods we know that

$$\hat{\beta} = (X'X)^{-1}X'y$$
$$\operatorname{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1},$$

  so we can always calculate the correlation among estimates.
- JMP, Design-Expert, and the `daewr` package in R will all make nice heatmaps of the correlations.

Classical Response Surface Methodology: Plackett-Burman Designs

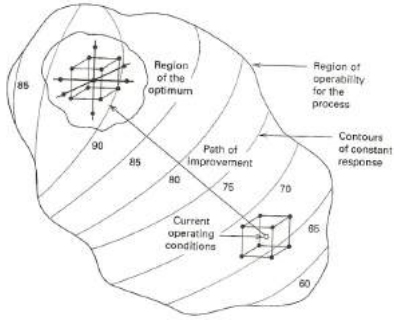


Correlation heat map of 2^{5-3}_{III} design and an 11-factor Plackett-Burman design

Classical Response Surface Methodology: Basic Strategy

1. Use a screening design (fractional factorial of resolution *III* or *IV*, Plackett-Burman, ...) to determine active factors
2. Use the effect estimates to move in the direction of steepest ascent
   - Include some center points in design to detect curvature
3. When curvature is detected, switch to designs sufficient for 2nd order models

Steps 1 and 2 or steps 2 and 3 can sometimes be combined.



Classical Response Surface Methodology: Screening

Three principles underlie the methods:

- *Hierarchical Ordering Principle* Lower order effects are more likely to be important than higher order effects
- *Effect Sparsity, or Pareto, Principle* The number of important effects is generally small
- *Effect Heredity Principle* For an interaction effect to be significant, (at least one of) its parent effects should be significant

Classical Response Surface Methodology: Screening

- When effects are not independent, normal & half-normal plots or the Lenth test are not really appropriate
  - Applies to non-regular designs such as Plackett-Burman, definitive screening, and others
- Instead need to build a model
  - Various approaches including enforcing hierarchical principle or Bayesian approaches with priors on main effects, 2-way interactions, etc
  - If there are not too many factors, can do all subsets regression and select using AIC_c, say

Classical Response Surface Methodology: 2nd Order Models

- With 2nd order designs, interested more in prediction than parameter estimation
- There are two classic 2nd order designs for k factors
- Central composite designs
  - require $2^{k-p} + 2k +$ number center points design points
  - typically use 5 levels; can use 3
  - there are also "small central composite designs"
- Box-Behnken designs
  - use 3 levels
  - do not use corners of experimental region
- (3^k designs used less often because of the size)

Classical Response Surface Methodology: 2nd Order Models

Number of runs

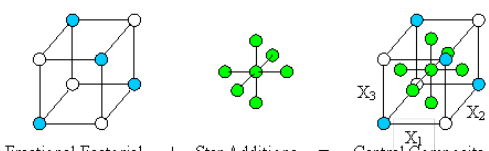
k	Central Composite Design	Box Behnken	Small Central Composite Design
3	10 or $14 + n_c$	$12 + n_c$	
4	16 or $24 + n_c$	$24 + n_c$	$19 + n_c$
5	26 or $32 + n_c$	$40 + n_c$	21 or $22 + n_c$

Classical Response Surface Methodology: 2nd Order Models

Central Composite Designs

Built out of

- A factorial part: 2^k or 2^{k-p}_{III} or 2^{k-p}_{V}
- Axial (aka "star") points, $2k$ of them, with all but one factor set to 0
- several points at the center



Classical Response Surface Methodology: 2nd Order Models

The radius, α, of the axis points is flexible

- $\alpha = 1$ is known as "face-centered" and so requires only 3 levels per factor
- $\alpha = 2^{(k-p)/4}$, makes the design "rotatable"—the variance of prediction depends only on the distance from the (coded) origin
- α can be used to make block effects orthogonal . . .

Classical Response Surface Methodology: 2nd Order Models

Blocking with central composite designs

1. If needed, block the factorial component, but include c_f center runs
2. One block for the axial points plus c_a center points
3. With the right choice of α, all the block effects will be orthogonal to the parameter estimates

$$\alpha = \sqrt{\frac{2^{k-p}(2k + c_a)}{2(2^{k-p} + c_f)}}.$$

- For example, with $k = 3$, $p = 0$, and $c_a = c_f = 2$, $\alpha \approx 1.79$

Classical Response Surface Methodology: 2nd Order Models

Small (Central) Composite Designs

- Replace the factorial component of a CCD with a partially aliased design such as Plackett-Burman
- Same idea works using 2^{k-p}_{III}
- The axial points resolve aliasing of 2-way interactions with main effects

Classical Response Surface Methodology: 2nd Order Models

Central composite designs

- Can combine with an existing factorial design to give a blocked 2nd order design
  - Advantageous for sequential experimentation
- Allow for 3 or 5 levels per factor
- Can be made (near) rotatable
- Flexible

Classical Response Surface Methodology: 2nd Order Models

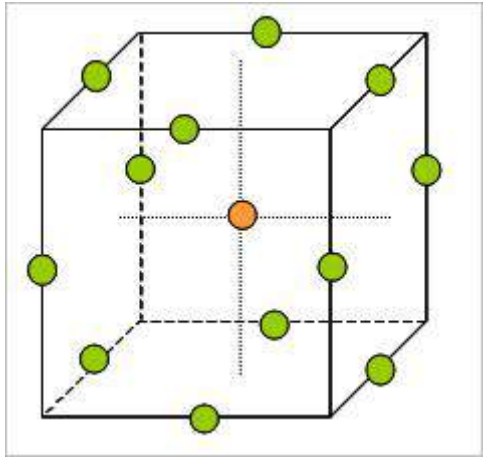
Box-Behnken designs

- Incomplete 3^k factorials
- Are (close to) rotatable—variance of prediction depends only on distance from center
- Formed by combining two-level factorial designs with incomplete block designs
  - E.g., $k = 3$,

A	B	C
$*$	$*$	
$*$		$*$
	$*$	$*$

- For each line, replace $*$ with a 2^2 factorial and put 0 in for blank cells

Classical Response Surface Methodology: 2nd Order Models



Note that there are no design points on the vertices.

Classical Response Surface Methodology: 2nd Order Models

Box-Behnken designs

- Have the advantage (and disadvantage) of using 3 levels per factor
- They do not extend to the vertices of the cube
  - May be an experimentation advantage
  - Don't think of as a cubic design—think of as a spherical design
- Are rotatable or near-rotatable
- Can sometimes block—3 blocks when $k = 3$ or 2 blocks when $k = 4$

Classical Response Surface Methodology: Canonical Analysis

- After the regression model has been estimated, the response can be studied with a contour plot or lattice plots of contours.
- Canonical analysis gives a direct estimate of the stationay point and the nearby behavior by writing the response surface in the form (e.g., when $k = 3$)

$$y - y_s = \lambda_1 w_1^2 + \lambda_2 w_2^2 + \lambda_3 w_3^2,$$

Classical Response Surface Methodology: Canonical Analysis

If we write

$$\hat{y} = \hat{\beta}_0 + x'\hat{\beta} + x'\hat{B}x,$$

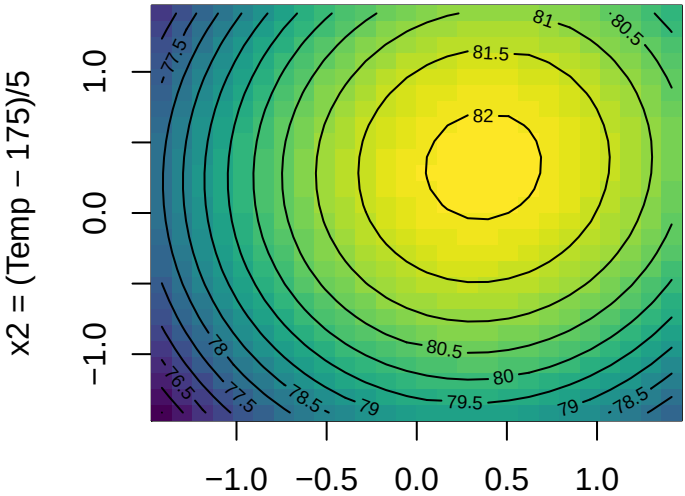
- To find the stationary point x_s, differentiate $\hat{y}$, set to zero, and solve

$$X_s = -\hat{B}^{-1}\hat{\beta}/2.$$

- The eigenvalues of $\hat{B}$ indicate the behavior at the stationary point.
- *We almost always want to not just know where an optimum is but how the surface behaves nearby.*

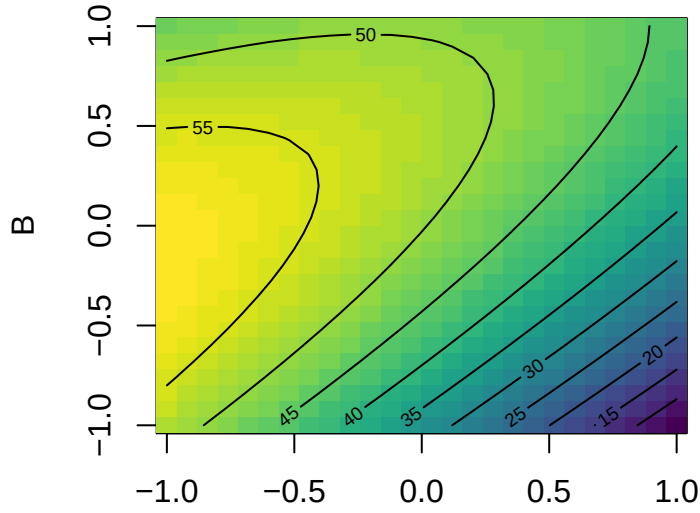
Classical Response Surface Methodology: 2nd Order Models

Myers, Montgomery, and Anderson-Cook, Example 7.6



Classical Response Surface Methodology: 2nd Order Models

Myers, Montgomery, and Anderson-Cook, Table 6.2



Classical Response Surface Methodology: Texts

1. Box [2009] and Box and Draper [2007] are classics
2. Myers et al. [2016] is the dominant textbook
3. Anderson and Whitcomb [2016] gives a light-footed introduction
4. del Castillo [2010] includes Bayesian approaches and computer experiments
5. Wu and Hamada [2011] gives a more mathematical treatment of topics

Optimal Design

Optimal design theory—and its implementations—are critical when classical designs are inappropriate

- Constrained design space
- Need a different number of runs than in classical designs
- Combination of mixture and process variables
- Split plot features
- Nonstandard linear models (e.g., cubic terms)
- Augmenting an existing design (particulary one with failed runs)

Optimal Design

Writing X as the design matrix,

- $\hat{\beta} = (X'X)^{-1}X'y$
- $\mathrm{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$
- $\hat{y}_x = x\hat{\beta}$
- $\mathrm{var}(\hat{y}_x) = \sigma^2 x(X'X)^{-1}x'$

The design problem:

- selecting rows of X
- such that the design defined by X in some defined sense, optimal
- Just need sensible criteria and algorithms

Optimal Design

Smith (1918) proposed "minimize the maximum variance of any predicted value"

$$\min_{x_i, i=1,\ldots,n} \max_{x \in \mathcal{X}} \operatorname{var}(\hat{y}_x).$$

Now called global, or *G-optimality*.

Optimal Design

Wald (1943) put the emphasis on the quality of the parameter estimates:

$$\max_{x_i, i=1,\ldots,n} |X'X|.$$

- Now called *D-optimality*
- Determinant is the product of the eigenvalues,
  - inversely proportional to the product of the axes of the confidence ellipsoid around $\hat{\beta}$
  - so maximizing $|X'X| \Leftrightarrow$ minimizing the volume of the confidence ellipsoid

The Kiefer-Wolfowitz General Equivalence Theorem established that D and G optimal designs are equivalent, under certain conditions.

Optimal Design

There are many other design criteria.

- I (or IV) optimal may be next most important
  - Minimizes the average variance of prediction over the region of interest
  - $\int_R x'(X'X)^{-1}x\,dx = \mathrm{Trace}\left((X'X)^{-1}M\right)$, where $M = \int_R xx'\,dx$
  - Over a cuboidal region minimizes the average variance of the parameter estimates

Optimal Design

There are many other design criteria.

- I (or IV) optimal
- Bayesian D-optimal: like D-optimal
  - gives relative weights to the importance of the model terms
  - considers $|X'X + K|$, where K is a diagonal matrix
    - 0 for terms of greatest interest
    - 16, say, for terms of secondary interest
- Bayesian I-optimal: like I-optimal,
  - gives relative weights to the importance of the model terms
  - $\mathrm{Trace}\left((X'X + K)^1 M\right)$

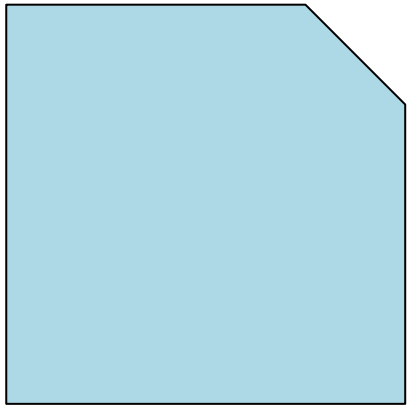
Optimal Design

- Use a design criterion that corresponds to your design problem
- Often
  - Estimation: D-optimality
  - Prediction: I-optimality

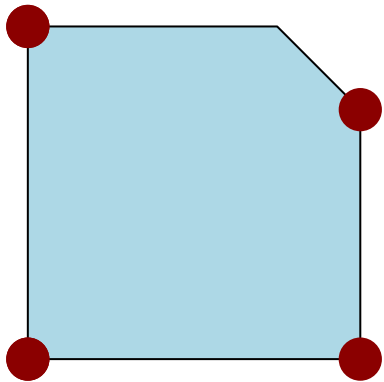
Optimal Design

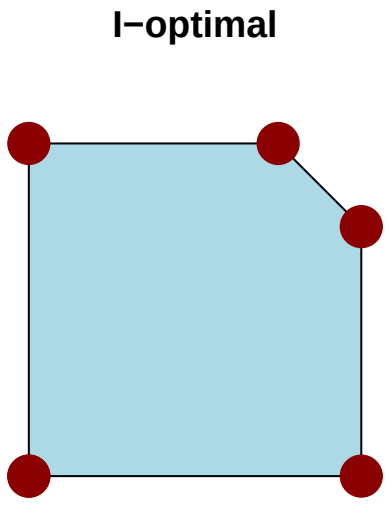
A toy example

First order model in a constrained region, $n = 6$

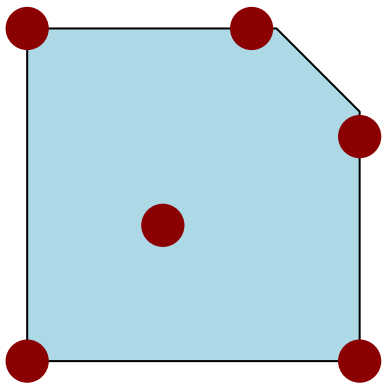


D−optimal

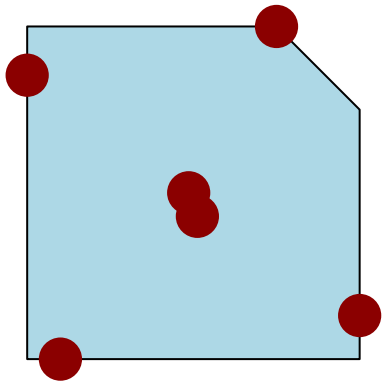



I−optimal

Bayesian D−optimal (2nd order terms if poss



Bayesian I−optimal (2nd order terms if poss



Optimal Design

- D- and I- optimal designs push points out to the edges
- Very model dependent
- To robustify the,
  - Previously, would use a higher order model or "D-optimal onion design"
  - Bayesian D-optimal and I-optimal designs target same objective more elegantly

Algorithms:

- Vertex exchange
  - Start with a random design, try to improve design point by point in the design matrix
- Coordinate exchange
  - Start with a random design, try to improve design element by element in the design matrix
  - Typically faster and more flexible
- Designs incorporate some randomness
  - Can get stuck in local optima
  - Use many restarts

Optimal Design

- JMP and SAS (`proc optex`) implement D-, I-, Bayesian D-, and Bayesian I-optimality
- Design-Ease implements D- and I-optimality
- R has `AlgDesign` package for D and I-optimality and some specialized ones

Goos and Jones [2011] is an engaging introduction to optimal design via case studies.

Definitive Screening Designs

Jones and Nachtsheim [Jones and Nachtsheim, 2011, 2013, 2016] introduced the class of "definitive screening designs"

- For m factors, they require $2m + 1$ runs
- Use 3 levels per factor
- Main effects are orthogonal
  - to themselves
  - to 2-way interactions
  - to quadratic effects
- 2-way interactions
  - not completely confounded among themselves or quadratic effects
- quadratic effects are
  - estimable
  - not completely confounded with 2-way interactions
- With 6 through 16 factors, all 2nd order models with ≤ 3 factors can be estimated
- With 17 through 20 factors, all 2nd order models with ≤ 4 factors can be estimated

Definitive Screening Designs

The design is built out of foldover pairs and a center run. The exact coordinates for each pair is chosen using optimal design criterion.

TABLE 1. General Design Structure for m Factors

Foldover pair	Run (i)	Factor levels				
		$x_{i,1}$	$x_{i,2}$	$x_{i,3}$	$\cdots$	$x_{i,m}$
1	1	0	± 1	± 1	$\cdots$	± 1
	2	0	∓ 1	∓ 1	$\cdots$	∓ 1
2	3	± 1	0	± 1	$\cdots$	± 1
	4	∓ 1	0	∓ 1	$\cdots$	∓ 1
3	5	± 1	± 1	0	$\cdots$	± 1
	6	∓ 1	∓ 1	0	$\cdots$	∓ 1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
m	$2m-1$	± 1	± 1	± 1	$\cdots$	0
	$2m$	∓ 1	∓ 1	∓ 1	$\cdots$	0
Centerpoint	$2m+1$	0	0	0	$\cdots$	0

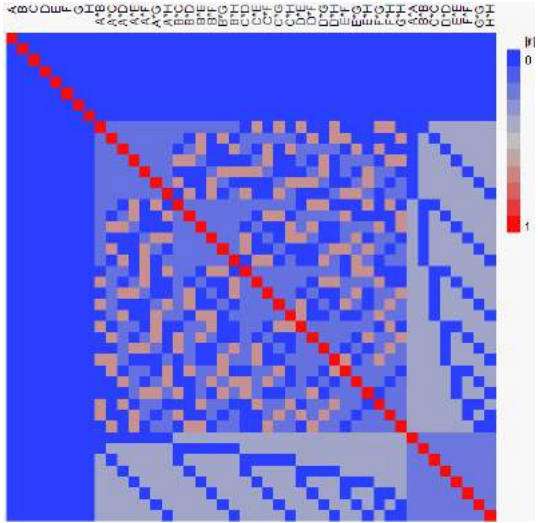
Definitive Screening Designs

This is an example of a 8-factor, 17-run design

	A	B	C	D	E	F	G	H
1	0	1	1	1	1	1	1	1
2	0	-1	-1	-1	-1	-1	-1	-1
3	1	0	-1	-1	1	-1	1	1
4	-1	0	1	1	-1	1	-1	-1
5	1	1	0	-1	-1	1	-1	1
6	-1	-1	0	1	1	-1	1	-1
7	1	1	1	0	-1	-1	1	-1
8	-1	-1	-1	0	1	1	-1	1
9	1	-1	1	1	0	-1	-1	1
10	-1	1	-1	-1	0	1	1	-1
11	1	1	-1	1	1	0	-1	-1
12	-1	-1	1	-1	-1	0	1	1
13	1	-1	1	-1	1	1	0	-1
14	-1	1	-1	1	-1	-1	0	1
15	1	-1	-1	1	-1	1	1	0
16	-1	1	1	-1	1	-1	-1	0
17	0	0	0	0	0	0	0	0

Definitive Screening Designs

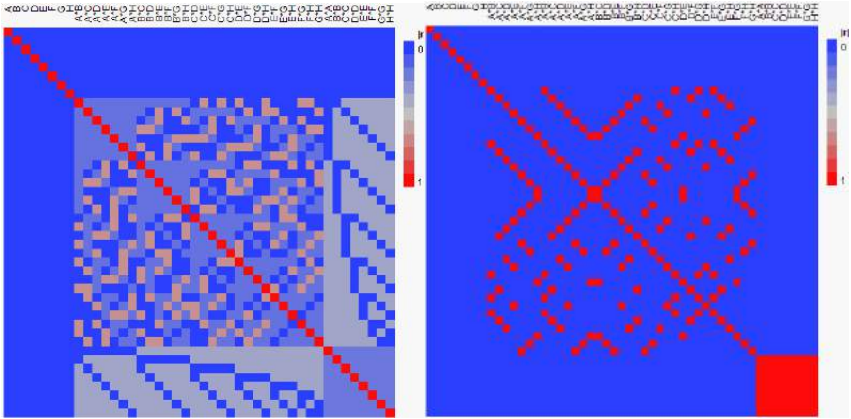
The heat map shows the correlation of the parameter estimates



from https://community.jmp.com/t5/JMP-Blog/Introducing-definitive-screening-designs/ba-p/30075

Definitive Screening Designs

Compare this to the parameter esimate correlation of a 2^{8-4}_{IV} design with 1 center point



from https://community.jmp.com/t5/JMP-Blog/Introducing-definitive-screening-designs/ba-p/30075

Definitive Screening Designs

- Definitive screening designs have been extended to incorporate flexible blocking and a few 2-value categorical variables
- Implemented in JMP and Design-Expert (and partially in R package `daewr`)
- Typically analyzed using
  - forward stepwise regression on a full 2nd order model
  - using the AIC_c criterion
  - sometimes with a hierarchy restriction
- Recommended only for at least 4 factors
- Can increase power by designing for more factors and dropping columns
  - e.g., for 4 factors, build a 6-factor DSD and drop the last 2 columns

Definitive Screening Designs

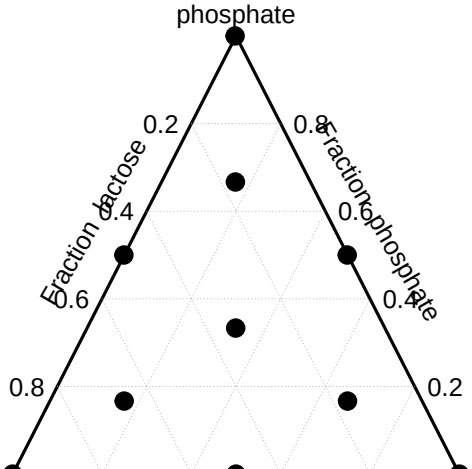
Not a fix-all. Not effective . . .

- For many categorical variables or categorical variables at more than 2 levels
- When you expect cubic effects
- When there are constraints on the design region
- When some factors are mixture ingredients
- When the design is run as a split-plot

Note: Christopher Nachtsheim is speaking on definitive screening designs on Wednesday at 8:30

Experimental Design for Mixtures: Classic Approach

Mixture experiments differ from others in that the components must add to one. We typically visualize the experimental region as a simplex.
A simple example from Waaler et al. [1991], via Eriksson et al. [1998].

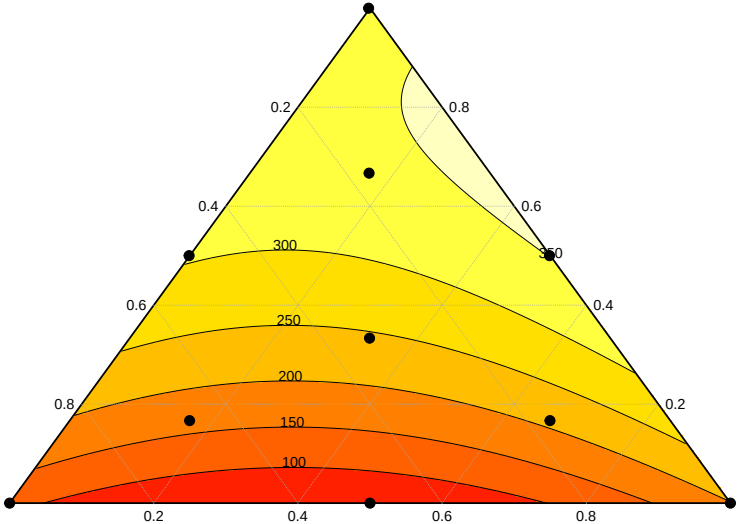


Experimental Design for Mixtures: Classic Approach

Scheffé Canonical Models

- Linear $E(y) = \sum \beta_i x_i$
- Quadratic: $E(y) = \sum \beta_i x_i + \sum \sum_{i<j} x_i x_j$
- Special Cubic: $E(y) = \sum \beta_i x_i + \sum \sum_{i<j} x_i x_j + \sum \sum \sum_{i<j<k} x_i x_j x_k$
- β_j represents the expected response from a pure mixture with $x_j = 1$
- $\sum \beta_i x_i$ term is called the linear blending term
- Quadratic terms are called nonlinear blending terms
- Estimates can be fit with OLS, but adjustment if needed for correct R square and standard errors of model coefficients

Experimental Design for Mixtures: Classic Approach

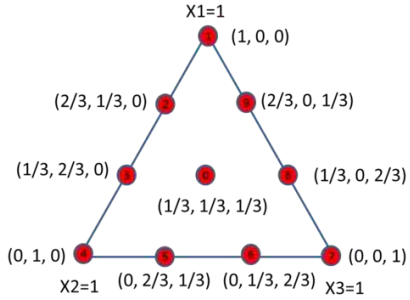



```
## [1] 0.04090909
```

Experimental Design for Mixtures: Classic Approach

Standard mixture designs 1

- Simplex-Lattice $\{q, m\}$ designs
  - Every possible combination of q components with $m + 1$ equally spaced values from 0 to 1, $x_i = 0, 1/m, 2/m, \ldots, 1$ for $i = 1, 2, ..., q$
  - Except centroid, all design points are on the boundaries of the simplex
  - often augmented with "axial check points" midway from vertices to centroid

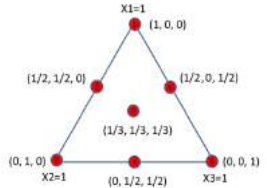


from http://reliawiki.org/index.php/Mixture_Design

Experimental Design for Mixtures: Classic Approach

Standard mixture designs 2

- Simplex-Centroid
  - q vertices of the form $(1, 0, \ldots, 0)$
  - all points of the form $(1/2, 1/2, 0, \ldots, 0)$
  - ...
  - all points of the form $(1/r, 1/r, \ldots, 0)$
  - 1 point of the form $(1/q, \ldots, 1/q)$.
- Except centroid, all design points are on the boundaries of the simplex
- often augmented with "axial check points" midway from vertices to centroid
- efficient for cubic models

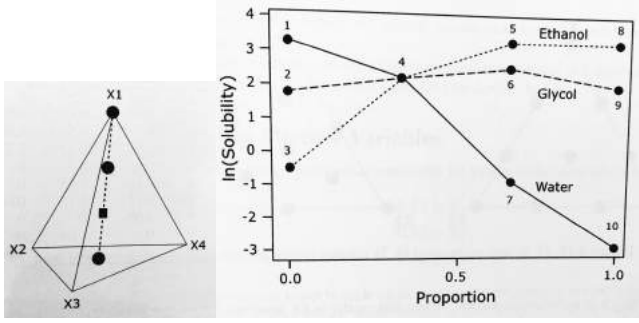


from http://reliawiki.org/index.php/Mixture_Design

Experimental Design for Mixtures: Classic Approach

Standard mixture designs 3

- Simplex Screening
  - Vertices
  - Centroid
  - Axial check blends



from @smith2005experimental

Experimental Design for Mixtures: Constraints

- It is not uncommon to have constraints in a mixture experimental space
- If every constraint has the form $x_i \geq L$, then the components can be rescaled to pseudo-components
- Interpretation of result can be helped by using a reference blend on Cox directions
  - Lines from vertex to reference blend
  - Proportion of other components is constant along that line
- Design is best handled via optimal design

Experimental Design for Mixtures

References:

- Cornell [2011]
- Smith [2005]

Analysis of Mixture Data Using Partial Least Squares

With many constraints and components,

- Scheffé model can be difficult to interpret
- Cox model (reference blend with Cox directions) is easier to interpret, but
- Estimation can be numerically unstable

Constraints lead to estimation problems

With contraints,

- the x_i increase in collinearity,
- $X'X$ becomes ill-conditioned, which
- leads to poor quality estimates

Including process (i.e., non-mixture) variables causes more complications.

Partial Least Squares

- *Partial least squares* (PLS) gives a single framework to analyze data from a mixture experiment, with or without process variables. Introduced by Kettaneh-Wold [1992] for mixture experiments
- Very popular in chemometrics, but less well known to statisticians
- A dimension reduction method
- Uses Y to identify orthogonal directions in X that are highly correlated with Y
- seeks factors that explain both response and predictor variation

Partial Least Squares

- The mth PLS direction $\hat{\phi}_m$ solves

$$\max_{\alpha} \mathrm{Cor}^2(y, X\alpha)\mathrm{Var}(X\alpha),$$

subject to $\|\alpha\| = 1$ and $\alpha \mathbf{S} \phi_{\mathbf{i}} = \mathbf{0}$, where $\mathbf{S}$ is the sample variance matrix of X.

- Compare to principal components analysis, which has criterion

$$\max_{\alpha} \mathrm{Var}(X\alpha),$$

subject to $\|\alpha\| = 1$ and $\alpha \mathbf{S} \phi_{\mathbf{i}} = \mathbf{0}$.

- Note that if X is orthogonal to start, then PLS is the same as ordinary least squares.
- Cross-validation is used to decide on how many dimensions to use.

Partial Least Squares

PLS is implemented in

- JMP
- `proc pls` in SAS
- `pls` package in R

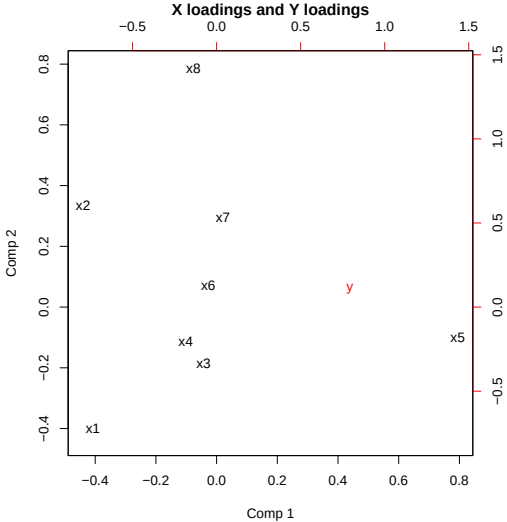
Partial Least Squares

A PLS Example (from Snee and Marquardt [1976] and Kettaneh-Wold [1992])
Eight components with the following constraints

- $0.10 \leq x_1 \leq 0.45$
- $0.05 \leq x_2 \leq 0.50$
- $0.00 \leq x_3 \leq 0.10$
- $0.00 \leq x_4 \leq 0.10$
- $0.10 \leq x_5 \leq 0.60$
- $0.05 \leq x_6 \leq 0.20$
- $0.00 \leq x_7 \leq 0.05$
- $0.00 \leq x_8 \leq 0.05$

A 20 run design, selected by D-optimality from 182 extreme vertices, plus 4 centroid replicates, was run.

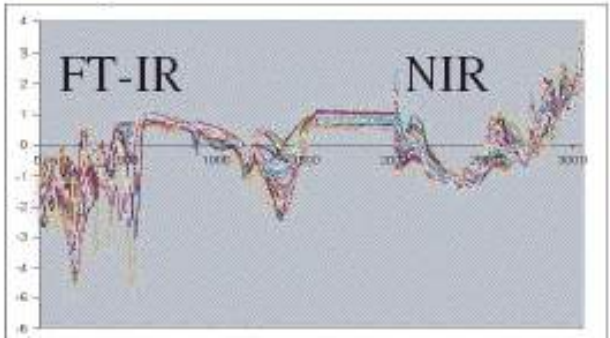
Partial Least Squares



As component 1 explains 97% of the variance, pretty clearly factors x_1, x_2, and x_5 have by far the largest effects. Kettaneh-Wold [1992]

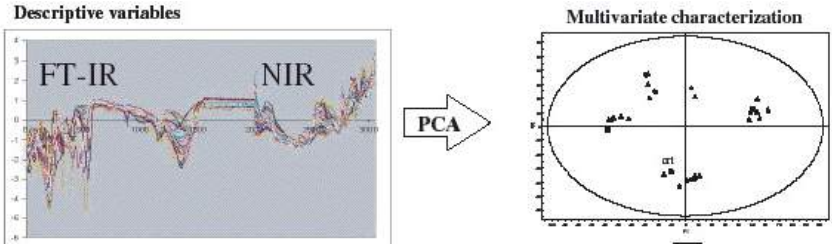
Chemometrics-style "Multivariate Design"

- May be the case that factors that can be manipulated are not the variables that directly affect the system (e.g., changing an amino acid changes many characteristics)
- May be too many possible factors (e.g., 400 pairs of amino acids, or many possible excipients from many suppliers)
- May also be too many variables, such as from spectroscopy



Chemometrics-style "Multivariate Design"

- As a first step, apply principal components analysis (PCA), reducing many variables to a few dimensions
- 2 approaches:
  1. Treat the PCA space as the set of factors, and choose design points from it corresponding to classic designs
  2. Treat the actual projected points in the design space as candidate points and apply, say, D-optimality



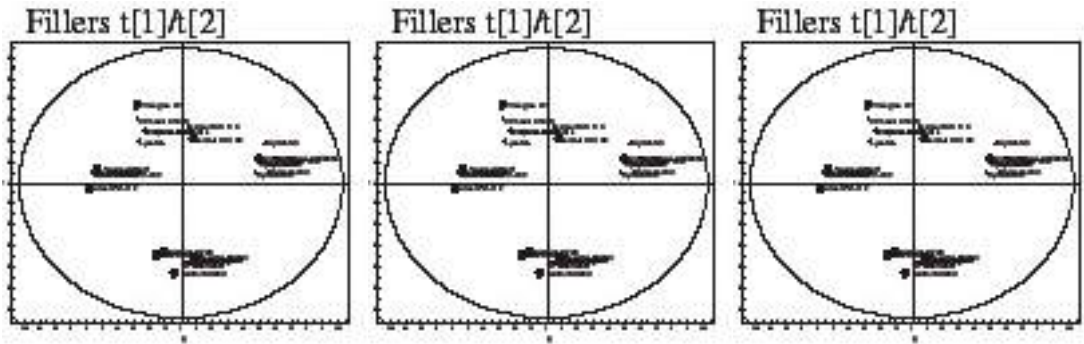
From @gabrielsson2003multivariate

Chemometrics-style "Multivariate Design"

An Example from Gabrielsson et al. [2003]
Screening of excipients to development of a new tablet formulation for direct compression. (They have a full set of papers describing various steps in the process.)

- 105 samples of 21 different excipients
- characterized by FT-IR and NIR spectroscopy
- divided into 3 classes (fillers, binders, and disintegrants), standardized separately
- (also samples of active pharmaeceutical ingredient (API) from three batches each from two manufacturers)
- PCA applied: 3 principal components extracted for each class

Chemometrics-style "Multivariate Design"

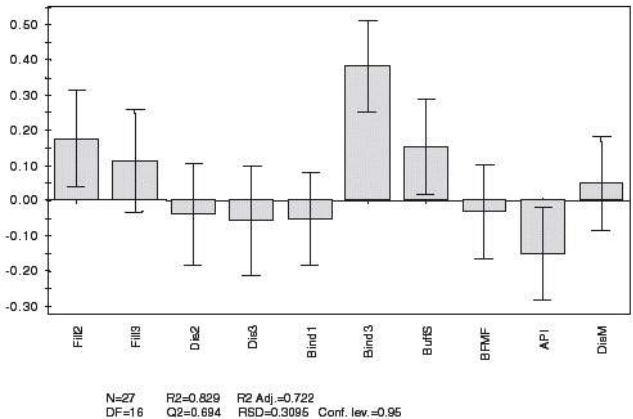


From @gabrielsson2003multivariate

Chemometrics-style "Multivariate Design"

- Total of 14 factors (3×3 principal components, plus API, plus some additional factors)
- Performed a 2_{IV}^{14-9} fractional factorial with 3 center points for 35 experimental runs
- Picked excipients corresponding to low and high values for each PCA component
- Measured 9 outcomes such as disintegration time and crushing strength
- Performed partial least squares with main effects only model for analysis

Chemometrics-style "Multivariate Design"



Factor	Abbreviation
Filler 1	Fill 1
Filler 2	Fill 2
Filler 3	Fill 3
Binder 1	Bind 1
Binder 2	Bind 2
Binder 3	Bind 3
Disint 1	Dis 1
Disint 2	Dis 2
Disint 3	Dis 3
Buffer type	BuffS
BFMF (ratio B/F)	BFMF
API score	API
Amount of buffer	BuffM
Amount of binder and filler	
Amount of disintegrant	DisM
Amount of MgSt and API	

Disintegration time

From @gabrielsson2003multivariate

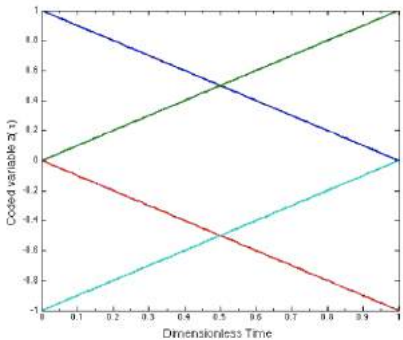
Chemometrics-style “Multivariate Design”

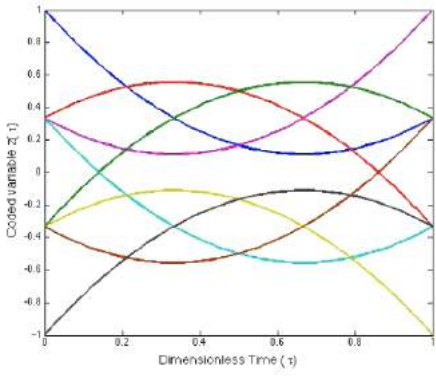
They separately considered - disintegration time - crushing strength - ejection force - disintegration type - hausner ratio - adhesion - compression force
It is still necessary to interpret the PCA components back to the original excipient samples.
Note: In the next paper in the series, Gabrielsson et al. [2006], they use D-optimality (using vertex exchange) to select the design points.

"Design of Dynamic Experiments"

An example of how we think of a design point can expand, from Georgakis [2013]

- Batch processes can involve time-varying inputs
- Expand these in terms of the first few Legendre polynomials
  - $P_0(\tau) = 1$, $P_1(\tau) = (-1 + 2\tau)/\sqrt{3}$, $P_2(\tau)(1 - 6\tau + 6\tau^2)/\sqrt{5}$, ...
  - The design points of a 2^2 factorial are then $\pm P_0(\tau) + \pm P_1(\tau)$





Design Space

In ICH Q8(R2), the *Design Space* is defined by

- The multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality.
- Working within the design space is not considered as a change.
- Movement out of the design space is considered to be a change and would normally initiate a regulatory postapproval change process

Design Space

This immediately brings to mind using designed experiments and statistical modeling. Two traditional approaches to multivariate response:

1. Overlapping mean response surfaces
2. Desirability functions

Multiple Responses: Overlapping mean response surfaces

Overlapping mean response surfaces is a straightforward approach

- Once an experimental region is set, the experimental design does not depend on the response
  - the design criteria are based on X, not y
- Following the experiment phase, build separate response surfaces for each response
- Define the design space as set as the values of factors such that the response surfaces are all within acceptable values

Multiple Responses: Overlapping mean response surfaces

Example 1: Response graphs for dissolution are depicted as a surface plot (Figure 1a) and a contour plot (Figure 1b). Parameters 1 and 2 are factors of a granulation operation that affect the dissolution rate of a tablet (e.g., excipient attribute, water amount, granule size.)

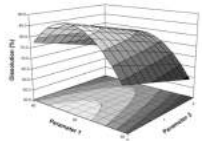


Figure 1a: Response surface plot of dissolution as a function of two parameters of a granulation operation. Dissolution above 80% is desired.

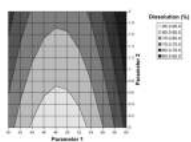


Figure 1b: Contour plot of dissolution from example 1a.

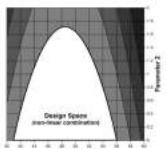


Figure 1c: Design space for granulation parameters, defined by a nonlinear combination of their ranges, that delivers satisfactory dissolution (i.e., >80%).

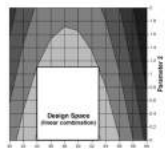


Figure 1d: Design space for granulation parameters, defined by a linear combination of their ranges, that delivers satisfactory dissolution (i.e., >80%).

Multiple Responses: Overlapping mean response surfaces

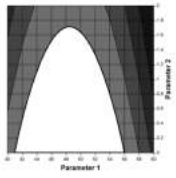


Figure 2a: Contour plot of dissolution as a function of Parameters 1 and 2.

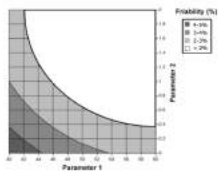


Figure 2b: Contour plot of friability as a function of Parameters 1 and 2.

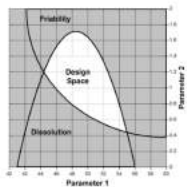


Figure 2c: Proposed design space, comprised of the overlap region of ranges for friability and/or dissolution.

Multiple Responses: Overlapping mean response surfaces

As an aside, in an optimization setting, the fitted response surfaces can be turned into a constrained optimization problem, e.g.
maximize y_1 subject to constraints on $y_2, \ldots, y_m$

Multiple Responses: Desirability functions

Idea:

- For each response, create an individual desirability function

$$0 \leq d_i \leq 1,$$

  setting $d_i = 0$ if the response is unacceptable.

Then maximize the overall desirability

$$D = (d_1 d_2 \dots d_m)^{1/m}.$$

Multiple Responses: Desirability functions

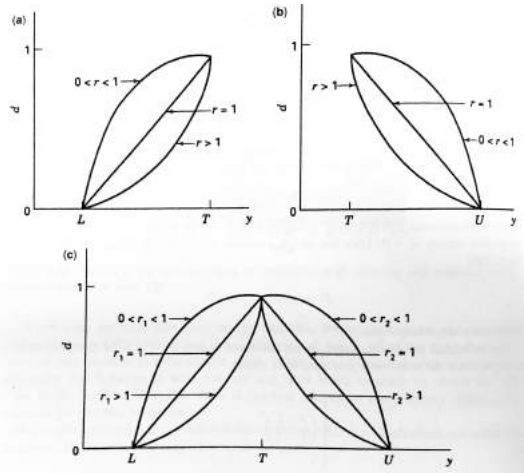


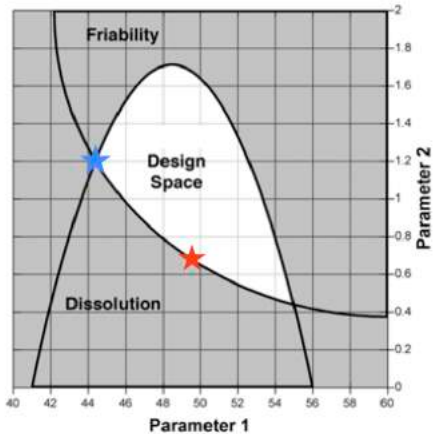
Figure 6.24 Individual desirability functions for simultaneous optimization y. (**a**) Objective (target) is to maximize y. (**b**) Objective (target) is to minimize y. (**c**) Objective is for y to be as close as possible to the target.

Multiple Responses: Desirability functions

- Not so simple to optimize, since the desirability function D is not differentiable
- Both Design-Expert and JMP have desirability function capabilities
- There is an R package `desirability`
- SAS requires (substantial) IML coding

Design Spaces: Problems with traditional multivariate response approach

- On the edge of a region defined by overlapping mean response surfaces, the probability that the responses exceed the limits is at best 50%.



Design Space: A frequentist approach

Chiao and Hamada [2001]'s approach:

- Write A for the specification region
  - Think of A as the region of acceptable values for the critical quality attributes (CQAs)
- Focus on $p(x) = P(Y \in A|x)$
- Assume Y has a multivariate normal distribution,

$$Y \sim N(\mu, \Sigma),$$

where $\mu = \mu(x)$ and $\Sigma(x)$ are functions of t x.

- For any A, μ, and Σ, finding $p(x) = P(Y \in A|x)$ corresponds to integration of the multivariate normal density

Design Space: A frequentist approach

Chiao and Hamada [2001] model

- μ_k in the usual way, as $x\alpha_k$,
  - α_k correspond to coefficients of a response surface model
- $\log(\sigma_k^2) = x\beta_k$
- $\tanh^{-1}(\rho_{kl}) = x\gamma_{kl}$ ($\tanh^{-1}(\rho_{kl})$ is equivalent to $1/2\text{logit}((\rho+1)/2)$)

Write $\Theta = (\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_m, \gamma_1, \ldots, \gamma_m)$

Design Space: A frequentist approach

To fit the model, there has to be replication at the design points so that the variance components can be modeled.

- Could be genuine replicates
- Could be across a small factorial of noise factors

Can then use sample estimates, $\hat{\mu}_k$, $\hat{\sigma}_k^2$, and $\hat{\rho}_{kl}$ as the responses to an unreplicated experiment
Note:

- $\log \hat{\sigma}^2$ is approximately normal with variance $2/(n-1)$ and
- $\tanh^{-1} \hat{\rho}$ is approximately normal with variance $1/(n-3)$

So significance tests can be used for determining the effects for those "responses"

Design Space: A frequentist approach

- Chiao and Hamada [2001] give 3 industrial examples of maximizing $p(x) = P(Y \in A | x, \text{data})$
- In practice,
  - $\hat{\mu}_k$, $\hat{\sigma}_k^2$, and $\hat{\rho}_{kl}$ can be modeled using standard software as responses to an unreplicated experiment
  - requires replication so that the covariance can be modeled
  - `mvtnorm` package in R can be used for the calculations
  - Could also use Monte Carlo simulation via draws from $N(\hat{\mu}, \hat{\Sigma})$
  - $P(Y \sim N(\hat{\mu}, \hat{\Sigma}) \in A$ can be calculated
    - over a grid and interpolated
    - using a computer experiment approach (e.g., a using a Latin hypercube followed by kriging, Fang et al. [2005])
- Note that uncertainty of model estimates is not incorporated

Design Space: Peterson's posterior predictive approach

Peterson [2004] proposed a Bayesian approach to incorporate both uncertainty and correlation.
Start with the model,

$$Y = Bz(x) + e, \qquad e \sim (N(0, \Sigma))$$

and consider how to determine $p(x) = P(Y \in A | x, \text{data})$

Design Space: Peterson's posterior predictive approach

To incorporate model parameter uncertainty,

- Peterson uses the usual noninformative joint prior for B and Σ
- proportional to $|\Sigma|^{-(p+1)/2}$

with the resulting posterior predictive distribution of . . .

Design Space: Peterson's posterior predictive approach

. . . $y|x, \text{data} \sim$ multivariate t-distribution, with

- mean $\hat{B}z(x)$,
- variance $D/\nu(1 + z(x)(Z(X)'Z(X)^{-1}z(x_0)'$, where $D = (Y - X\hat{B})'(Y - X\hat{B})$ is a scale matrix
- $\nu = n - p - q + 1$ degrees of freedom

Computation can be done

- `mvtnorm` package in R, using the Genz and Bretz algorithms
- Monte Carlo simulation, converting multivariate normal random variables and independent chi-square random variables to multivariate t

Design Space: Peterson's posterior predictive approach

Extension to Seemingly Unrelated Regression (SUR) Models, Peterson et al. [2009]

- No requirement that the regression model be the same for each outcome

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}$$

- Also uses multivariate t distribution instead of multivariate normal
  - allows for moderate outliers

Design Space: Peterson's posterior predictive approach

- Models for each response can be done separately using usual tools
- Residuals can be used to estimate multivariate error correlation structure
- Peterson et al. [2009] uses Gibbs sampling to simulate draws from the posterior predictive distribution of $\mathbf{Y}$ given a value of x

Design Space: Peterson's posterior predictive approach

"Pre-posterior analysis"
After design, experimentation, and analysis, the region of acceptable probability of conformance to specifications may be too small, or the probabilities too low.
Inherent in the process, or a function of the sample size?

- Impute additional data from the fitted model
- Redo the analysis with new data
- Examine the change in the probability of conformance
  - A sufficient change indicates more certainty in parameter estimates (i.e., bigger sample size) is needed
  - An insufficient change indicates that the issue lies with the process itself

Bayesian predictive approach

Advantages of the Bayesian predictive approach [from **?**]:

- It incorporates model parameter uncertainty
- It incorporates correlations among responses
- Informative priors can be used if available
- Direct calculation of a region of acceptable probability of conformance to specifications
- Allows noise variables to be added
- Allows "pre-posterior analysis"
- Very general and flexible framework

Bayesian predictive example 1

Lebrun et al. [2012] give a particularly complete example, in optimizing a spray-drying process for a pulmonary drug.
There were three critical process parameters:

Critical process parameters	Abbreviation	Low level	High level
Inlet temperature (°C)	IT	110.00	220.00
Spray flow rate (L/h)	SFR	439.00	1744.00
Feed rate (ml/min)	FR	2.50	7.50

Bayesian predictive example 1

and five critical quality attributes:

CQA	Specification
Yield	>80
Moisture	<1
Fraction [1–5] m	>60
Compressibility index	<15
Hausner ratio	<1.18

The latter two, measures of flowability, are derived from bulk and tapped density.

Bayesian predictive example 1

Lebrun et al. [2012] used a multivariate multiple linear regression model, with

$$Y = (\text{logit}(\text{Yield}), \text{logit}(\text{Fraction}), \log(\text{Moisture}), \log(\text{Bulk}), \log(\text{Tapped}))$$

and

$$Y = \beta_0 + \text{linear}(IT, FM, SFR) + \text{quadratic}(IT, FR) + \text{2-way interaction}(IT, SFR) + \text{3-way interactio}$$

with

$$E \sim N(0, \Sigma).$$

By specifying a particular non-standard model they are trying to prevent over- and under-fitting.

Bayesian predictive example 1

As above, by using the non-informative prior $p(B, \Sigma) = |\Sigma|^{-(p+1)/2}$ they arrive at predictive posterior density of

$$(Y|x_0, \text{data}) \sim T_p\left(\hat{B}x_0, D/\nu(1 + x_0(X'X)^{-1}x_0', \nu)\right),$$

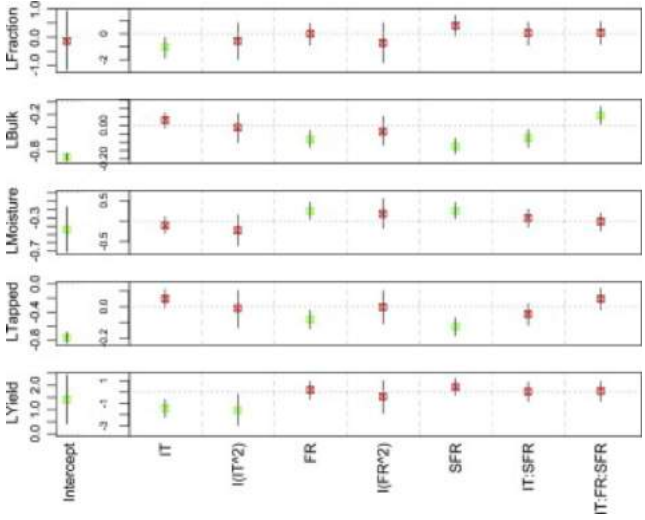
where

- $\hat{B}$ is the least square estimate of B
- $D = (Y - X\hat{B})'(Y - X\hat{B})$ is a scale matrix
- $\nu =$ number design pts $-$ number responses $-$ number factors $+ 1$

Bayesian predictive example 1

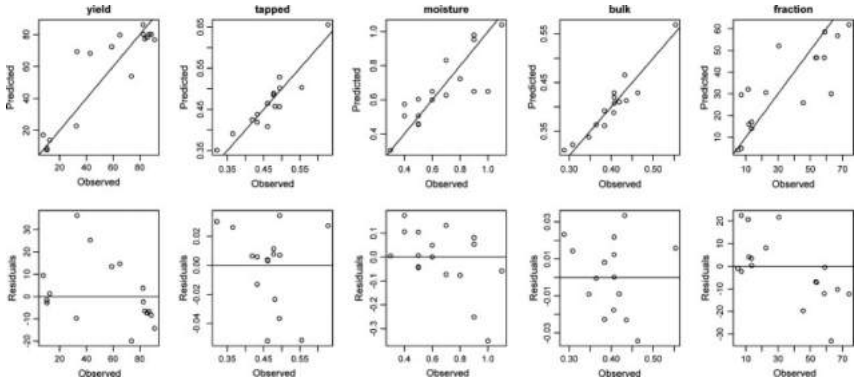
- For the design, they used a face-centered central composite design with 3 center points for $n = 17$.
- First, they use the marginal posterior density of parameters to get 95% credible intervals

Bayesian predictive example 1



Bayesian predictive example 1

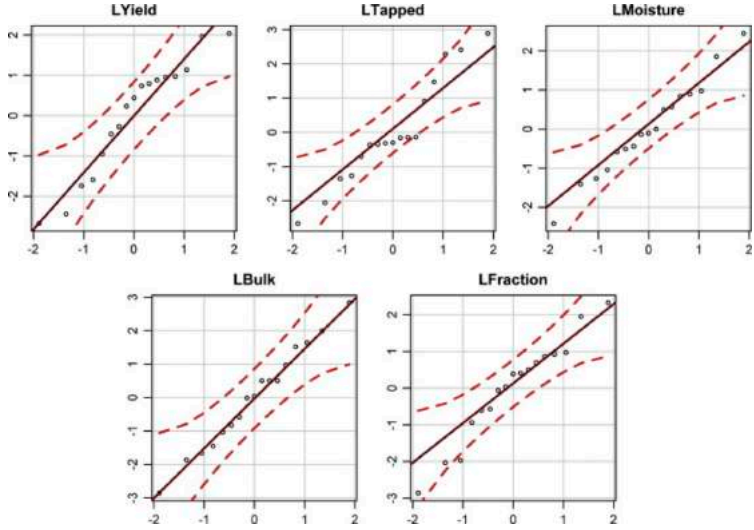
Next, they examine residuals



Note the amount of variability

Bayesian predictive example 1

... including normal QQ plots



Bayesian predictive example 1

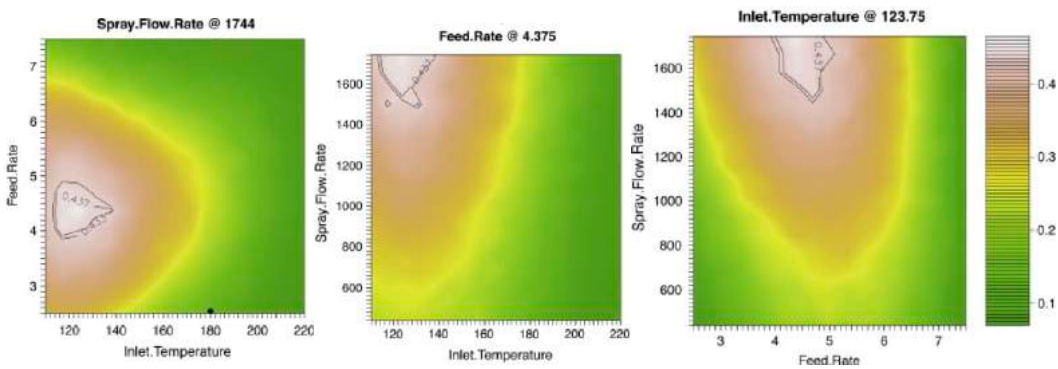
The estimated correlation matrix is a long way from the identity matrix.

```
## Warning: Missing column names filled in: 'X1' [1]
```

X1	LYield	LTapped	LMoisture	LBulk	LFraction
LYield	1				
Ltapped	−0.05	1.00			
Lmoisture	0.16	0.04	1.00		
LBulk	0.3	0.90	0.31	1.00	
LFraction	0.83	0.32	0.28	0.60	1

Bayesian predictive example 1

The resulting probability map



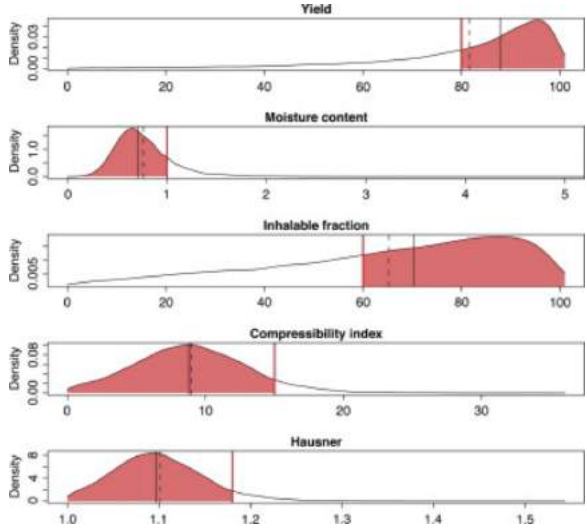
Bayesian predictive example 1

... and a resulting cuboid design space (for Prob > 0.437)

Critical process parameters	DS range
Feed rate (ml/min)	[4.2–4.8]
Spray flow rate (L/h)	[1614–1744]
Inlet temperature (°C)	[118–125]

Bayesian predictive example 1

They finish by examing the marginal distributions at the optimum



Bayesian predictive example 1

CQA	Lower 95%	Lower 75%	Median	Mean	Upper 75%	Upper 95%
Yield (%)	42.00	75.00	88.00	81.00	94.00	100
Moisture content (%)	0.26	0.57	0.71	0.76	0.89	1.31
Inhalable fraction (%)	17.00	49.00	70.00	65.00	85.00	100
Compressibility index	0.40	6.20	8.80	9.00	11.50	16.1
Hausner ratio	1.00	1.07	1.09	1.10	1.13	1.19

Bayesian predictive example 2

- Fu et al. [2016] give an example of a yeast fermentation process for producing a recombinant protein.
- 3 critical process parameters: pH, disolved oxygen, temperature
- 10 critical quality attributes

Broth Titer (g/L)	yHCP (ppm)	yDNA (pg/mL)	Impurity 6AA (%)	Purity by RP-HPLC (%)	Protease Level (U/µg)	cIEF Main Peak (%)	SEC Aggregate (%)	Purity by cGE (%)	Potency (%)
≥1.12	≤1.76E+06	≤4.38E+07	≤11.62	≥74.98	≤358.93	≥42.38	≤11	≥93.01	≥50 and ≤15

- Central composite design with 8 center points, augmented with 6 more
  - One axial point (DO) run farther out than others
- Used Peterson et al. [2009] approach

Bayesian predictive example 2: Overlapping contour plot

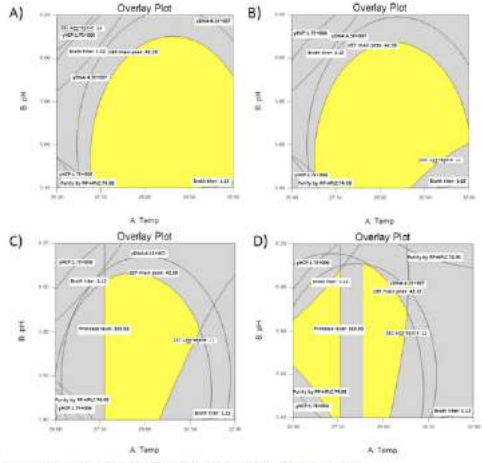
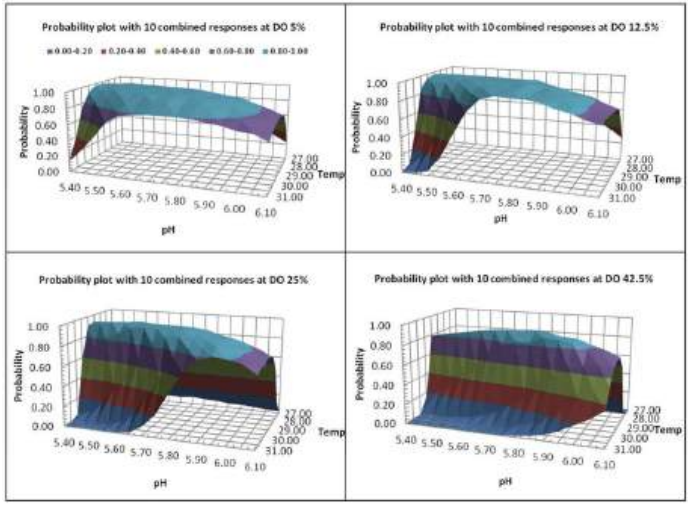


Figure 1. The overlay region comprised of all response attributes meeting their acceptance criteria.

The contour overlay plot was conducted and analyzed using the DESIGN EXPERT 7.1.1. The relations of three critical process parameters, temperature, pH and DO to six response attributes, broth titer, yHCY, yDNA, impurity 6AA, purity by HP-HPLC, protease level, cHEF main peak, HCC aggregate, purity by cCIE, and potency, meeting their acceptance criteria (Table 1) are shown at DO level of A) 5.0%; B) 12.5%; C) 25.0%; and D) 45.0%.

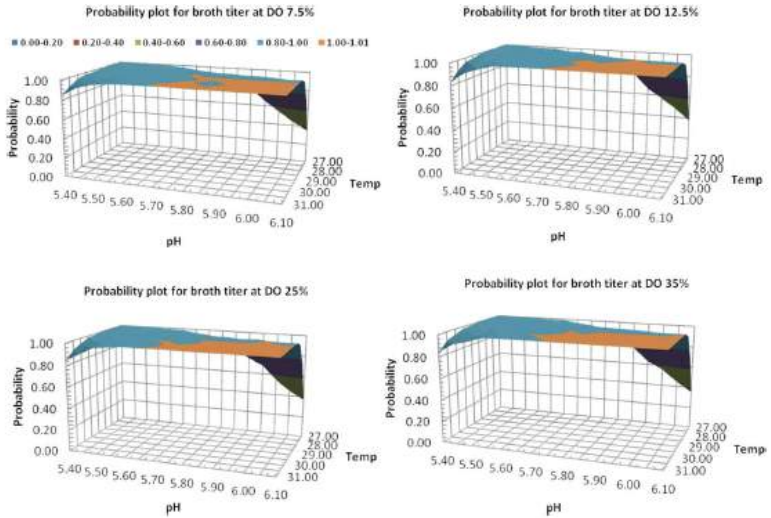
from @fu2016characterization

Bayesian predictive example 2: predictive probability plot



from @fu2016characterization

Bayesian predictive example 2: predictive probability plot of broth titer $\geq$ 1.12 g/L



from @fu2016characterization

Take-away

References

M.J. Anderson and P.J. Whitcomb. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments, Second Edition*. CRC Press, 2016. ISBN 9781315351728. URL https://books.google.com/books?id=WI2KDQAAQBAJ.

G.E.P. Box. *Statistics for Experimenters: Design, Innovation, and Discovery, Second Edition + JMP Version 6 Software Set*. Wiley, 2009. ISBN 9780470570913. URL https://books.google.com/books?id=WOQZPwAACAAJ.

G.E.P. Box and N.R. Draper. *Response Surfaces, Mixtures, and Ridge Analyses*. Wiley Series in Probability and Statistics. Wiley, 2007. ISBN 9780470072752. URL https://books.google.com/books?id=pTb9PgbQAvYC.

Chih-Hua Chiao and Michael Hamada. Analyzing experiments with correlated multiple responses. *Journal of Quality Technology*, 33(4):451, 2001.

John A Cornell. *Experiments with mixtures: designs, models, and the analysis of mixture data*, volume 895. John Wiley & Sons, 2011.

Enrique del Castillo. *Process optimization: a statistical approach*, volume 105. Springer Science & Business Media, 2010.

Lennart Eriksson, Erik Johansson, and Conny Wikström. Mixture design—design generation, pls analysis, and model usage. *Chemometrics and Intelligent Laboratory Systems*, 43(1):1–24, 1998.

Kai-Tai Fang, Runze Li, and Agus Sudjianto. *Design and modeling for computer experiments*. CRC Press, 2005.

Zhibiao Fu, Daniel Baker, Aili Cheng, Julie Leighton, Edward Appelbaum, and Juan Aon. Characterization of a saccharomyces cerevisiae fermentation process for production of a therapeutic recombinant protein using a multivariate bayesian approach. *Biotechnology progress*, 32(3):799–812, 2016.

Jon Gabrielsson, Nils-Olof Lindberg, Magnus Pålsson, Fredrik Nicklasson, Michael Sjöström, and Torbjörn Lundstedt. Multivariate methods in the development of a new tablet formulation. *Drug development and industrial pharmacy*, 29(10): 1053–1075, 2003.

Jon Gabrielsson, Michael Sjöström, Nils-Olof Lindberg, Ann-Christin Pihl, and Torbjörn Lundstedt. Multivariate methods in the development of a new tablet formulation: excipient mixtures and principal properties. *Drug development and industrial pharmacy*, 32(1):7–20, 2006.

Christos Georgakis. Design of dynamic experiments: A data-driven methodology for the optimization of time-varying processes. *Industrial & Engineering Chemistry Research*, 52(35):12369–12382, 2013.

Peter Goos and Bradley Jones. *Optimal design of experiments: a case study approach*. John Wiley & Sons, 2011.

Bradley Jones and Christopher J Nachtsheim. A class of three-level designs for definitive screening in the presence of second-order effects. *Journal of Quality Technology*, 43(1):1, 2011.

Bradley Jones and Christopher J Nachtsheim. Definitive screening designs with added two-level categorical factors. *Journal of Quality Technology*, 45(2):121, 2013.

Bradley Jones and Christopher J Nachtsheim. Blocking schemes for definitive screening designs. *Technometrics*, 58(1):74–83, 2016.

Nouna Kettaneh-Wold. Analysis of mixture data with partial least squares. *Chemometrics and Intelligent Laboratory Systems*, 14(1-3):57–69, 1992.

Pierre Lebrun, Fabrice Krier, Jérôme Mantanus, Holger Grohganz, Mingshi Yang, Eric Rozet, Bruno Boulanger, Brigitte Evrard, Jukka Rantanen, and Philippe Hubert. Design space approach in the optimization of the spray-drying process. *European Journal of Pharmaceutics and Biopharmaceutics*, 80(1):226–234, 2012.

R.H. Myers, D.C. Montgomery, and C.M. Anderson-Cook. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley Series in Probability and Statistics. Wiley, 2016. ISBN 9781118916018. URL `https://books.google.com/books?id=YFSzCgAAQBAJ`.

John J Peterson. A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology*, 36(2):139, 2004.

John J Peterson, Guillermo Miro-Quesada, and Enrique del Castillo. A bayesian reliability approach to multiple response optimization with seemingly unrelated regression models. *Quality Technology & Quantitative Management*, 6(4):353–369, 2009.

W.F. Smith. *Experimental Design for Formulation*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics, 2005. ISBN 9780898715804. URL `https://books.google.com/books?id=m8kEgk9X8U4C`.

Ronald D Snee and Donald W Marquardt. Screening concepts and designs for experiments with mixtures. *Technometrics*, 18(1):19–29, 1976.

PJ Waaler, C Graffner, and BW Müller. Optimization of a matrix tablet formulation using a mixture design. *Acta pharmaceutica nordica*, 4(1):9–16, 1991.

CF Jeff Wu and Michael S Hamada. *Experiments: planning, analysis, and optimization*, volume 552. John Wiley & Sons, 2011.