Statistical Methods for Establishing the Design Space and Control Within Design Space

H. Gregg Claycamp, MS, PhD
Office of New Animal Drug Evaluation
Center for Veterinary Medicine
FDA/OFVM

1

Disclaimers, etc.,

- The thoughts presented here are the author's and none of this material is intended to represent either policy or guidance of FDA, whether existing or contemplated.
- This is a presentation of research results and review of generalized principles for DOE.
- The literature cited is representative of a wealth of information available from research libraries. Some of this work and was done prior to the author joining the FDA.

Overview

- Intro to general design space and DOE theory from risk/stats POV
  - Design of Experiments
  - Models, models, and more models…
  - Types of designs: "what's the question?"
- Background to design space uncertainty
  - A thought experiment about design space uncertainties and risk
  - Bayesian Monte Carlo approaches for Multivariate Design Space

Typical Themes from Engineering Statistics*

1. Explore
2. Measure
3. Characterize

} Prepare; develop

4. Model
5. Improve

} Design of Experiments

6. Monitor
7. Compare
8. Reliability

} Maintain quality

*This a Chapter list in *NIST Engineering Statistics Handbook.*
IMHO--Design space concept of interest is Eng. Stat. derived.

Paths to the Design Space (DS)

- In ICH/FDA Q8 (R2) and others, the notion of the final design space is multivariate.
- Examples in GFI show univariate design space (process modeling) and combining results → multivariate view of the DS.
- DOEs are uni- or multi-variate in nature and are intended to create efficient, and direct solutions.
- Depending on prior knowledge about the processes, investigators might use process modeling for a detailed understanding.

How do we get the best coefficients?

- The notion is for predictability⇔controllability
- In general: $y = f(\vec{x}; \vec{\beta}) + \epsilon$
- Least-squares criterion
  - Majority of methods use L-S minimization for best $\vec{\beta}$ estimates
- Design of Experiments (DOE) principles
  - For a given set of x, $\vec{\beta}$ cannot be improved. But DOE principles are to select values of x that can target improvement of the estimates for a desired outcome.

Assumptions in (Process) DOE Modeling*

- The process is a statistical process
- The means of the random errors = 0
- The random errors have constant standard deviations
- The random errors follow a normal distribution
- The data are randomly sampled from the process
- The explanatory variables are observed without error

*e.g., *NIST/SEMATECH e-Handbook of Statistical Methods*

Design of Experiments (DOE)

- R.A. Fisher in 1926 and 1935, (*The Design of Experiments)*
  - Fisher statistical DOE a.k.a. “controlled experiments”
- G.E.P. Box, 1950s → Regression, factorial methods.
- DOE: one of several approaches to describing a design space.
- Box, *et al*: Problem-solving approach set at data collection stage to ensure valid and supportable engineering conclusions.
- Performed under constraints of minimal expenditure of experimental runs, time and money.

All good science begins with a question!

- Any high-quality scientific, analytical, or engineering exercise begins with well-formed questions and an objective.
- Prior knowledge helps experts formulate clear questions and testable hypotheses.
- Seldom is there a unique solution to a complex problem: solid first principles in design help to
  - Identify the type of problem and method needed
  - Define the goals—comparison, optimization, min/max, etc.
- DOE, as part of QbD, is a principled way to achieve solutions to specific engineering, chemistry and analysis problems.

Purposes/Types of Design of Experiments

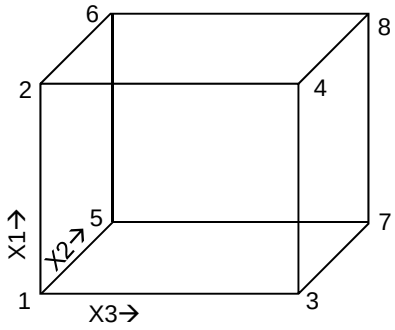
- Comparative
  - Effect of changing a parameter on the process
- Screening/characterizing
  - Process understanding—which variables are important?
- Modeling
  - Creating a good fit and accurate quantitative model
- Optimization
  - Determine best settings of factors (parameters) under the desired response

DOE Objectives→ Design Selection

- **Comparative objective**:
  - One or several factors under investigation,
  - Goal to make a conclusion about one a-priori important factor
  - Question of interest is whether or not that factor is "significant"
- **Screening objective (main effect designs)**:
  - To select or *screen out* the few important main effects from the many possible.
- **Response Surface (method) objective**:
  - Estimate interactions and/or quadratic effects→ shape of the response surface.
  - Find improved or optimal process settings
  - Troubleshoot process problems and weak points
  - Make a product or process more *robust*
- **Optimizing responses when factors are proportions of a mixture**
  - Mixture design.
- **Optimal fitting of a regression model objective**:
  - Desire "good" model parameter estimates (i.e., unbiased and minimum variance).
  - If you want to model a response as a mathematical function (either known or empirical) of a few continuous factors.

Design for success?

- Example: By covering enough parameter space and interactions to create a robust model; e.g.,
  - 2^3 Factorial design
  - 3 Factors (X1,X2,X3)
  - 2 Levels
  - No *blocking* shown

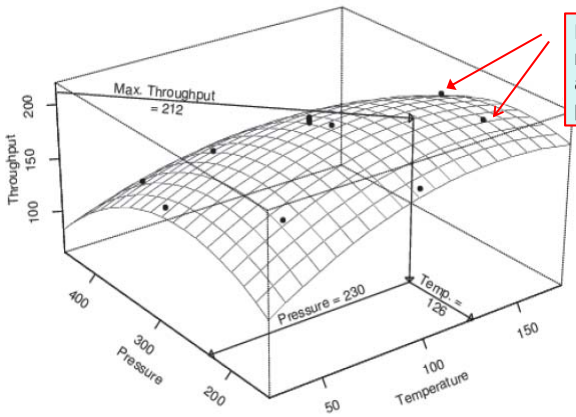


Simple 2^3 Factorial Design, continued

- The simple design can "cover" three factors and two levels/factor in eight runs.
- But, designs become quite complex with more factors and the need to cover, say, mid-points *and* end points.
- Purpose of the DOE (four general purpose) creates a variety of designs—many do not have every possible combination of factor levels

Factor Levels			
Run No.	X1	X2	X3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Examples cont'd, Optimization Designs



Points on a surface. Chose model points to help find an optimum (stationary point) for the process.

*From *NIST/SEMATECH e-Handbook of Statistical Methods*

Which Design Depends on #Factors & Objectives

Number of Factors	Comparative Objective	Screening Objective	Response Surface Objective
1	Completely randomized		
2 - 4	Randomized block design	Full or fractional factorial	Central composite or Box-Behnken
5 or more	Randomized block design	Fractional factorial or Plackett-Burman	Screening to reduce number of factors

- Each design calls for fixed and targeted levels of the factors: a "designed' experiment.
- Given the numbers of factors, levels, and blocking levels possible is a design, fortunately software is used to assign levels.

Regression Tools Dominate the Modeling

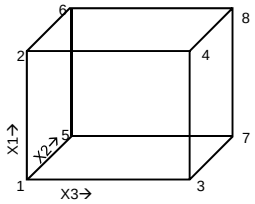
- Perhaps most DOE and design space problems are linear combinations… (linear in the parameters)

$$f(\vec{x};\vec{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots$$
$$f(\vec{x};\vec{\beta}) = \beta_0 + \beta_1 x + \beta_{11} x^2$$
$$f(x;\vec{\beta}) = \beta_0 + \beta_1 \ln(x)$$

are each linear with regard to parameters.

- Models can be univariate (above) or multivariate in which the independent term is also vector: $\vec{y}$ systems of equations
- Many software choices for implementing models.

Models, cont'd



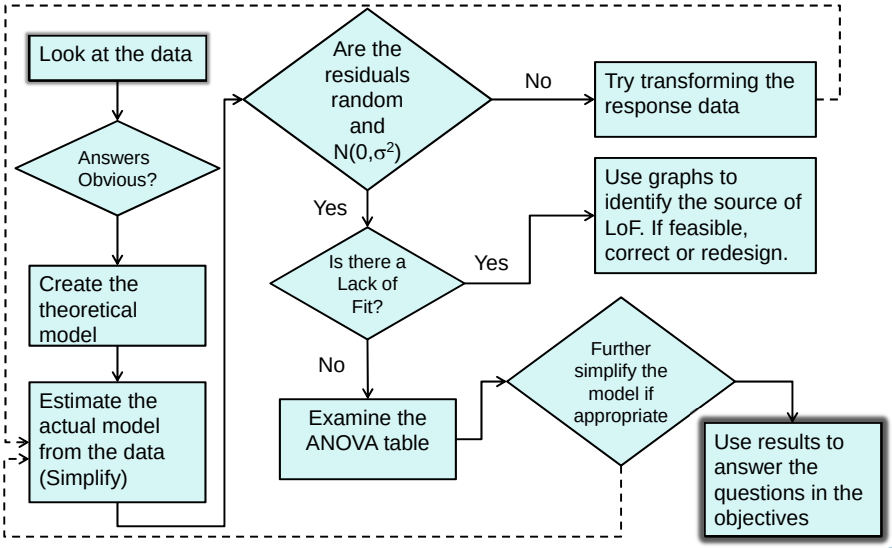
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

- In place of the 3-way interactions, might have a quadratic component (curvature):

$\ldots \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \ldots$

- In general, only significant terms ($p < 0.05$) are kept for the final model and analysis.

DOE Analysis Steps (After NIST)

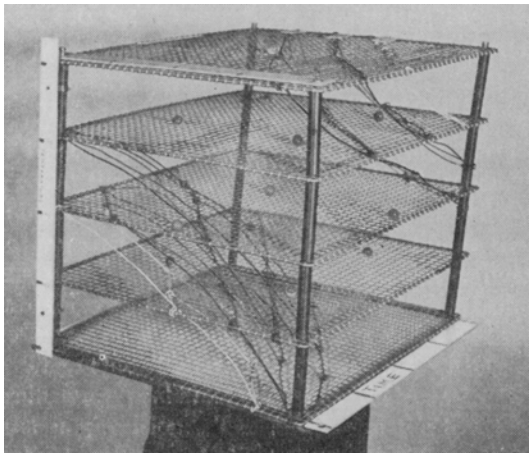


G.E.P. Box's Design Space in 3-D!

We have come a long way!

- Figure showing response surface modeling and results.

GEP Box, Biometrics 10:16-60 (1954).



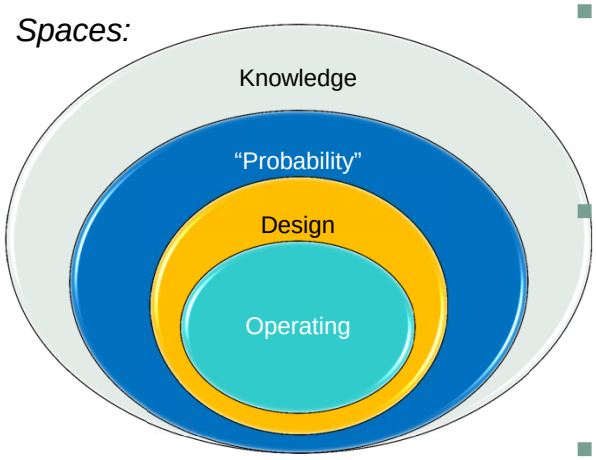
Other "spaces" exit?

Certainly!

(Particular personal interest: "probability" and "risk" spaces.)

Often neglected or assumed inconsequential…

Spaces:



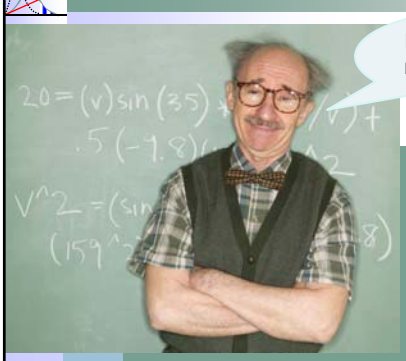
- "Probability"—for example—referring to the uncertainty about the design space.
- At a given 100(1-α) level, a confidence region might be:
  - > Design
  - < Design
  - Or = Design
- Does it matter??

Risk & Uncertainty Definitions for Today's Talk*

- *Risk = exposure to a chance of loss (--something of value). (Generalized scientific notion of risk)*
- *Risk = probability of harm x severity of that harm. (FDA/ICH Guidelines)*

- A Risk Analyst's notion of uncertainty:
  - Uncertainty is both *aleatory* and *epistemic* in nature
    - Aleatory ~ variation about an expected value
    - Epistemic ~ due to lack of knowledge; degree of belief.

**Many specific definitions exist in different contexts.*



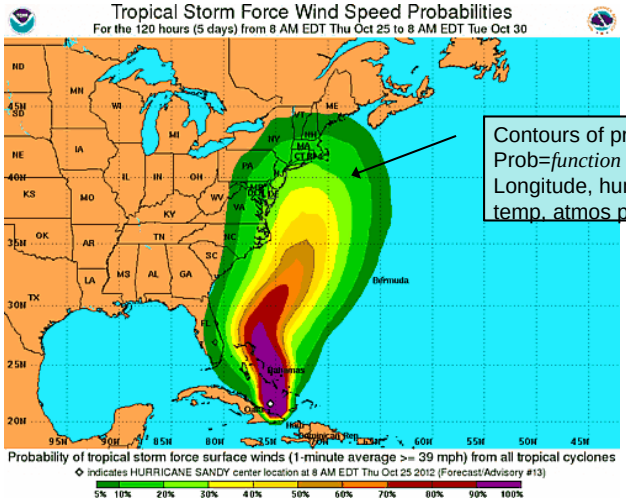
“All models are wrong but some are useful.”*

*Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer, and G. N. Wilkinson, (eds.) *Robustness in Statistics*. New York: Academic Press.

23

Risk Analysis ◊ Decision Analysis

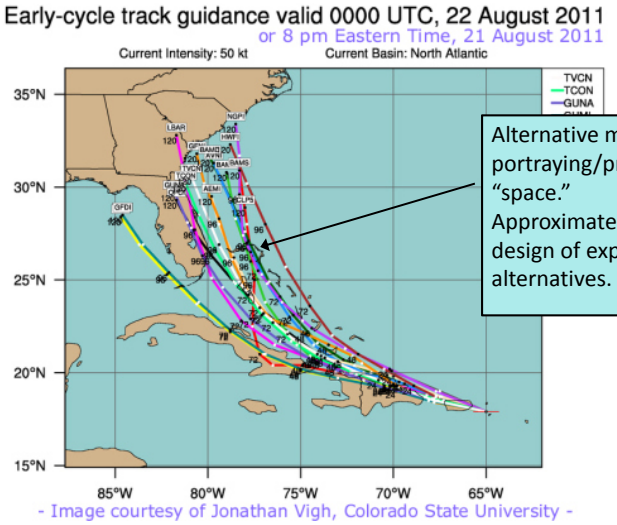
A “Risk-based” Thought Experiment: “Design Space” for a Hurricane



Contours of probability: Prob=*function of* (time, Latitude, Longitude, humidity, surface temp, atmos press,…)

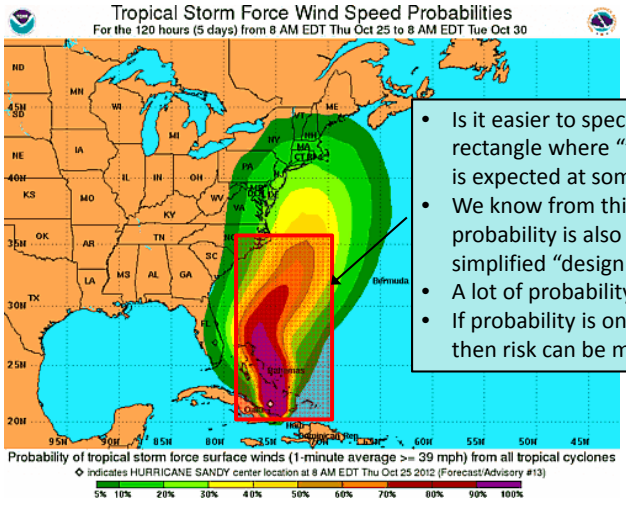
Claycamp, 2003-2013

Model Uncertainty



Alternative models available for portraying/predicting the same "space."
Approximately analogous to design of experiment (DOE) alternatives.

Shaping a Design Space?



- Is it easier to specify a (LAT x LON) rectangle where "trouble" (wind speed) is expected at some probability?
- We know from this illustration that probability is also important outside the simplified "design space."
- A lot of probability is missed!
- If probability is only partially modeled, then risk can be more uncertain.

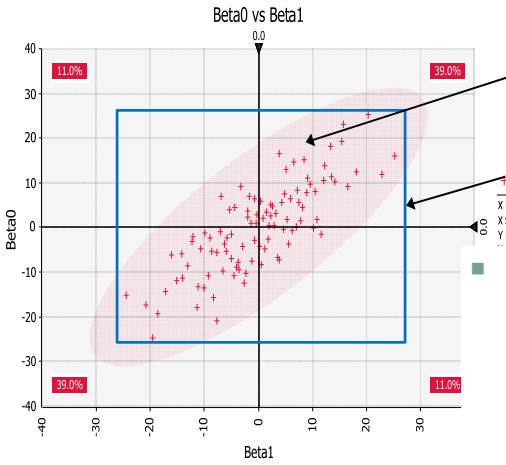
Thought experiment, continued.

- In the hurricane "design space" (DS), risk increases proportionally to probability of storm exposure.
- Pharmaceutical DS risk is proportional to [1-prob(DS)].
- Pharmaceutical design space:
  - Minimize risk to product/process quality
  - Possibly optimize process
  - Possibly include other factors in creating the space, such as economics, up/downstream factors.
- How can risk and & uncertainty be described and predicted?

Thought experiment, continued.

- If the model is correct, then we expect to operate in the design space with a precision estimate from the regression and analysis of variance.
- The model, using OLS, GLM, or FGLS provides the averages—the "central tendency" model.
- BUT, things go wrong and things vary!!
- The uncertainty measures, as confidence intervals, confidence regions, tell us more about the expected range, given the uncertainty.

Confidence Ellipses and Intervals ($y = \beta_0 + \beta_1 x$)*



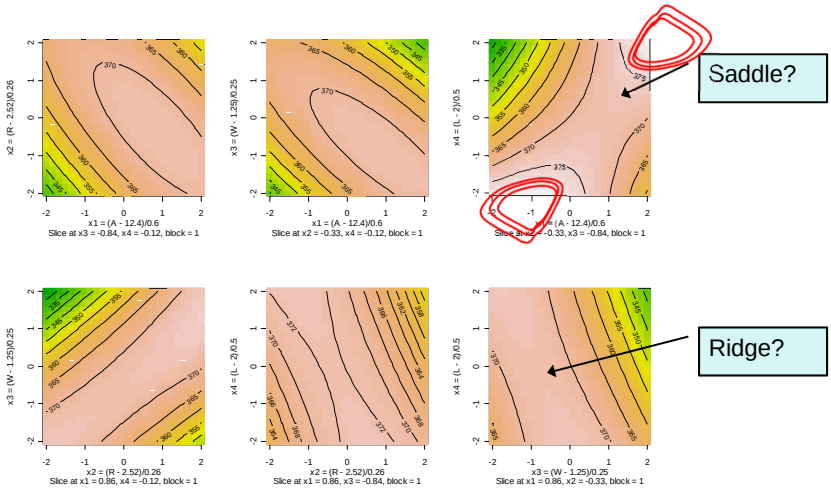
100(1-α) Joint confidence region for Beta0 and Beta1 (ellipse)

Bonferroni-type confidence intervals (conservative)

- Depending on the model for calculating intervals and regions, coverage of the actual data varies.

*(Simulated data using 100 pairs of ~*N*(0,100) random numbers; correlated draws; $\rho = 0.8$.)

Example contour views for a response surface

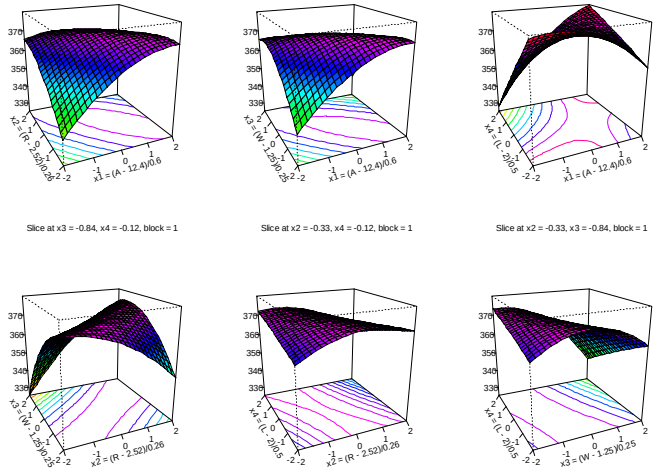


Saddle?

Ridge?

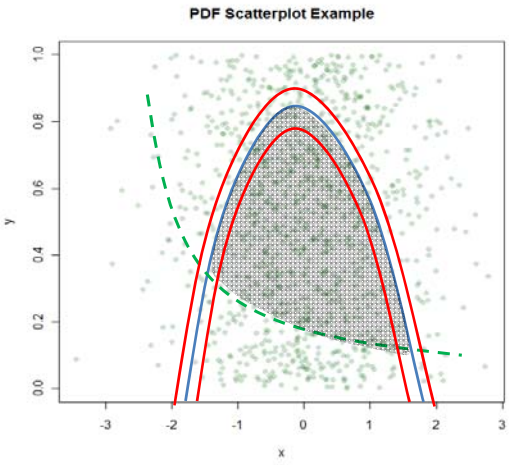
After Lenth, RV (2010) Surface plots in the rsm package

Surface→ Contour Perspectives. Uncertainties?



Joint p(Design Space) from Univariate DSs?

- A proposed design space (DS) represented by the shaded area – the DS dictates "Less than the *blue* function AND greater than the *green* (dashed line) function".
- Note that expected random errors find many (x,y) pairs outside of the DS.
- By sampling the posterior distribution –the DS—the values of parameters for *blue* and *green* can be found that more closely satisfy the design space (shaded region only).



One Possible Solution: Bayesian Approach

- In a perfect world, the process parameters (θ_k) are modeled from a DOE then are *controlled* during manufacturing.

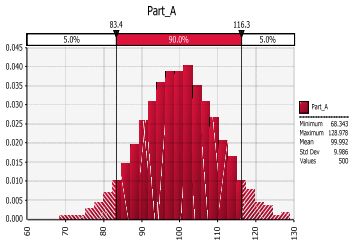
1. Reality—
   - *The means (perfect* control of a complex process is not possible.
   - "Noise" variables occur. These may be statistically described, but remain random events under an *assumed* probability density.
   - Noise variables might have actual distributions differing from the assumptions.
   - In multivariate setting, n^{th}-order correlation possible.

Why Use a Bayesian Approach? (2)

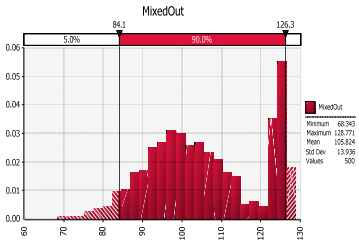
2) Uncertainty in the final design space improves predictions about the process risk control.
   - Classical Fisherian methods: Intersecting contours and joint uncertainty distributions are often calculated under *assumptions* of normality of the underlying errors.
   - Monte Carlo (MC) sampling approach can explore the posterior distribution in any form.
   - MC and Markov Chain MC (MCMC) methods can be computationally efficient optimizers when encountering a complex multivariate design.

- Real Time Release
  - Bayesian methods "update" information

Example: One-Dimensional Illustration…



Assumed model posterior.



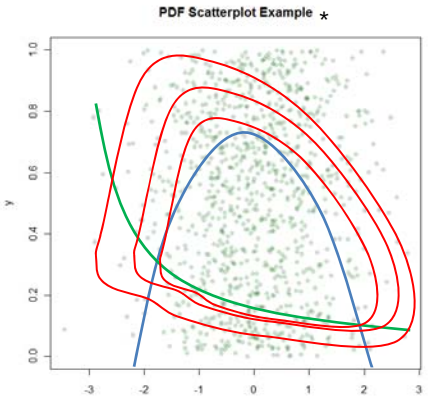
Bayes model posterior.

- Left: an assumed normal distribution used to predict uncertainty in the posterior predictive (quality) parameter.
- Right: Bayesian MC posterior shows a local minimum, suggests a mixture of posteriors distributions possible.

(Note: simulated data).

Joint Confidence Regions?

- Computational tools exist for predicting the joint confidenc in a multi-variate environment.
- Markov-chain Monte Carlo sampling is one established approach.



**Example only: simulated data and arbitrary contours*

When does it matter to know confidence regions?

- ICH Q9 principle to use the right risk management for the job!

One notion: the DS is already in a statistically robust region(?)

Estimated Confidence Regions are sufficient?

2nd notion: As a source of uncertainty and risk in the DS, the model and data uncertainties *should be assessed and then judged significant or not compared to other uncertainties.*

*Example only: simulated data and arbitrary contours

Examples: Bayesian Strategy w/ Simulation

- We are seeking design space parameters such that the probability that outcomes, $\boldsymbol{Y}$, are in the acceptable criteria (design space), $\boldsymbol{A}$, *given* the current set of parameters $\boldsymbol{x}$ and the *data,* is greater than or equal to a threshold reliability, $\boldsymbol{R}$.*

$$\left\{\mathbf{x} : \Pr(\mathbf{Y} \in A \mid \mathbf{x}, \text{data}) \geq R\right\}$$

*See Peterson et al. (2008; 2009)

- The samples also create joint posterior probability density for and for the marginal values.

Seemingly Unrelated Regression (SUR)

- Say we have factors *A,B,C* that predict the CQAs we are interested in, *X, Y, Z*.
- We are interested in a simultaneous solution for CQAs within lower (LL) and upper limits (UL) , e.g., $LL_X \le x \le UL_X$.
- Typical problem: A system of simultaneous equations
- Through univariate experiments or a DOE, we find that *A,B,C* and the interactions, *AB,AC,BC*, do not appear in every equation:

$$x = A + B + C + AB + AC + BC + \varepsilon_x$$
$$y = A + \cdots + C + \cdots + AC + \cdots + \varepsilon_y$$
$$z = A + \cdots + C + \cdots + \cdots + \cdots + \varepsilon_z$$

where ε_i are random error terms.

In Matrix Form and using math notation…

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$$

Where

$$X_1 = (1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3,)'$$
$$X_2 = (1, x_1, x_3, x_1x_3\ )'$$
$$X_3 = (1, x_1, x_3)'$$

i.e., each of the significant terms in the *y* equations will be expressed as a diagonal element of the matrix.

β_i = the vector of coefficients for each of the equations

ε_i = the error terms vector

In a compact, matrix notation:

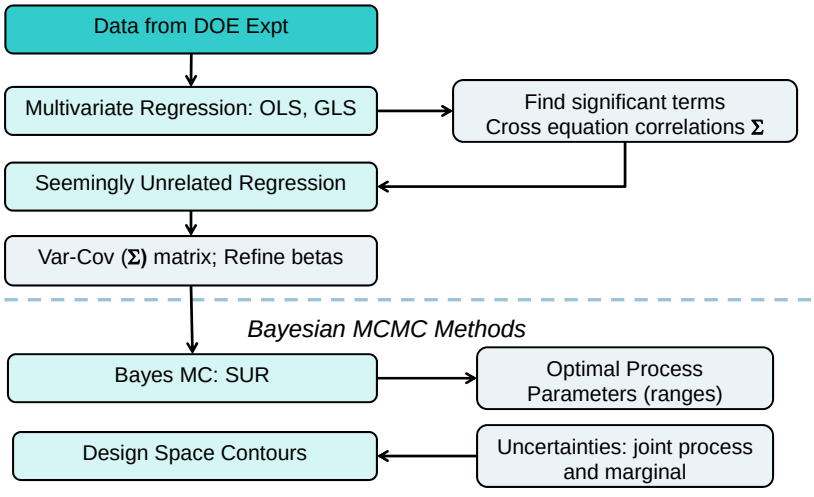
$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$$

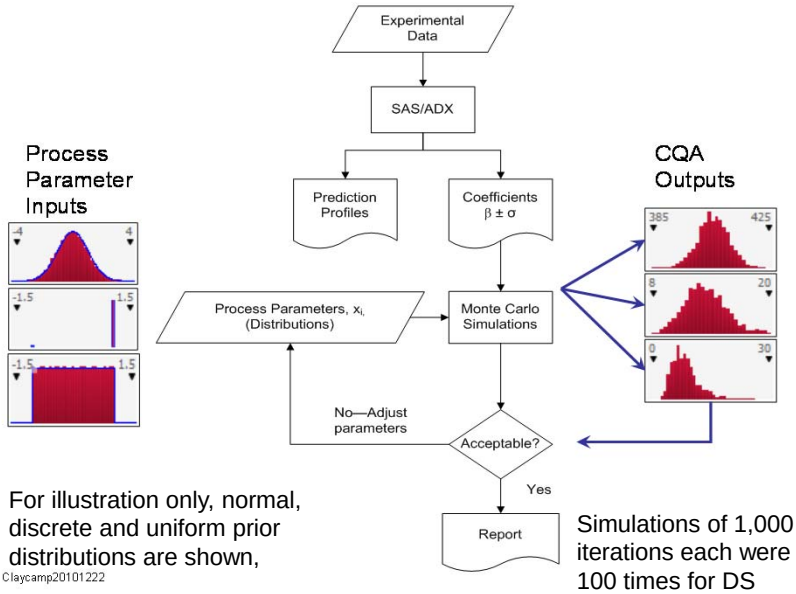
A SUR and Gibbs Sampled MCMC

$$\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{I})$$

- The key to SUR is that e are correlated between equations through the cross-model covariance, $\boldsymbol{\Sigma}$.
- Gibbs sampling calls for sampling from conditional posterior pdfs for $(\beta|\Sigma)$ and $(\Sigma|\beta)$.
- The link between the two distributions are through residuals $(y - X\beta)$

SUR – Gibbs Sampling Approach for DS





For illustration only, normal, discrete and uniform prior distributions are shown,

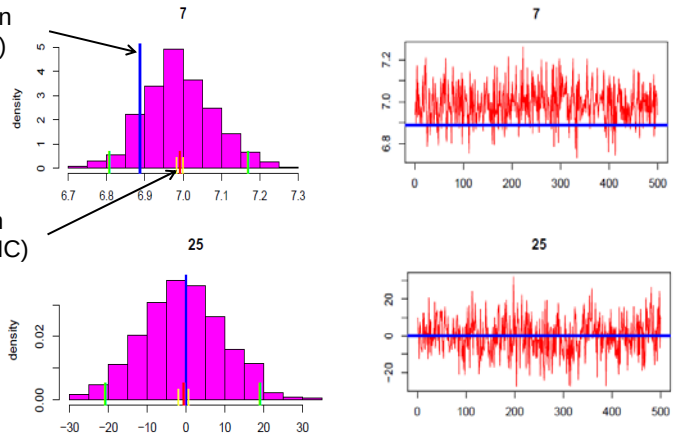
Simulations of 1,000 iterations each were run 100 times for DS estimation.

Examples for MCMC diagnostics

Samples: N=5000 keep=10th

Proposed mean estimate (SUR)

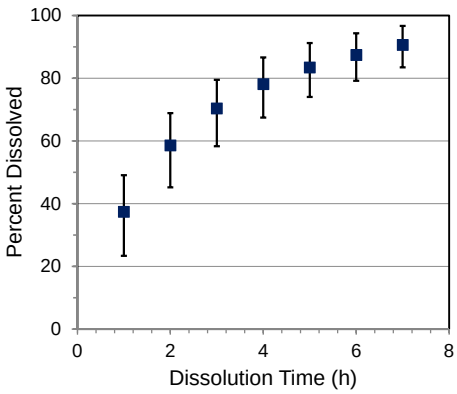
Posterior mean estimate (MCMC)



Current Observations

- Classical ANOVA and Response Surface method (RSM) parameters focused the design space closely. The Bayesian-SUR sampling trimmed only a small portion of the range of 1 out of 4 process parameters.
- Using f2 ≥ 50 as the acceptability criterion for samples,
  - Any combination of the seven dissolution time points could be used.
  - We looked as 3-way combinations (as in FDA GFI), and all 7.
- The Bayesian method readily fit all 7 time points without assumptions about the underlying dissolution model (e.g, Weibull, etc.)
- Posterior prob ~74% chance (depending on sampling parameters) that all 7 will meet the DS dissolution criteria at once.

Posterior Samples for which f2 ≥ 50



- Simulations for which f2 ≥ 50 are shown and the 2.5 and 97.5 percentiles on the samples.
- Note: For and 3-time-point strategy, a single simulation can explore all 35 combinations of 3 time points, some of which follow the GFI criteria.

Posterior Cross-Model Correlation (Σ): SUR

Correlation Matrix (rsurGibbs)							
	Disso1	Disso2	Disso3	Disso4	Disso5	Disso6	Disso7
Disso1	1	0.91	0.74	0.56	0.44	0.35	0.28
Disso2		1	0.93	0.83	0.74	0.67	0.61
Disso3			1	0.97	0.92	0.87	0.81
Disso4				1	0.98	0.96	0.91
Disso5					1	0.99	0.96
Disso6						1	0.98
Disso7							1

- Near-neighbor correlations (near-diagonal) ≅ SUR-estimates.
- Distant correlations (e.g., [Disso1,Disso7] decreased—the model is focused.

Posterior Cross-Model Correlation (Σ): RSM

Correlation Matrix (rsm)							
	Disso1	Disso2	Disso3	Disso4	Disso5	Disso6	Disso7
Disso1	1	0.954	0.894	0.808	0.729	0.677	0.578
Disso2		1	0.967	0.923	0.870	0.830	0.731
Disso3			1	0.982	0.951	0.919	0.831
Disso4				1	0.988	0.968	0.889
Disso5					1	0.991	0.939
Disso6						1	0.965
Disso7							1

- Higher correlation likely due to better match in design
- The betas in RSM show the variability across time in dissolution

Conclusions

- Design of Experiments (DOE)– long history from Fisher (sample stats) to G.E.P. Box (engineering stats) and beyond.
- Targeting design → Highly effective strategies for process improvement and design.
- Under-appreciated:
  - Assumption in designs including "normal" errors and the "centering" effect of RSM.
  - Uncertainties are "ball-parked" and are assumption-rich, using various frequentist statistic methods.
- Ball-park estimates might suffice for well-controlled DS.
  - The story is still unknown overall; but, worth at least a glance in a risk-based decision making approach!

Acknowledgments

- Raafat Fahmy
- Bhavesh Kothari
- Steve Hoag
- Ahmed Ashour

FDA Guidance / ICH Guidelines

- Guidance for Industry, "Q8(R2) Pharmaceutical Development" (Nov 2009) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073507.pdf
- Guidance for Industry, "Q9 Quality Risk Management" (June 2006) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073511.pdf
- Guidance for Industry, "Pharmaceutical Quality Systems" (April 2009) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073517.pdf
- Guidance for Industry, "Q8, Q9, and Q10 Questions and Answers" (May 2010) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM210822.pdf

References

- Altan, S., Bergum, J., Pfahler, L., Senderak, E., Sethuraman, S., Erland Vukovinsky, K. (2010) Statistical considerations in design space development (Part I of III) *Pharmaceutical Technology,* **34:** 66-70.
- Boulanger, B., Dewé, W., Gilbert, A., Govaerts, B., Maumy-Bertand, M. (2007), Risk management for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and the in-study validation phases. *Chemometrics Intel Lab Sys.* **86**:198-207.
- del Castillo, E. and Cahya, S. (2001) A tool for computing confidence reigions on the stationary point of a response surface. *Amer Statistician* **55**(4): 358-365.
- Draper, N.R., and Guttman, I. (1995) Intervals versus regions. *JRSS Series D*, *The Statistician* **44**(3):399-403.
- *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, Accessed May 5, 2013.
- Peterson, J.J., Miró-Quesada, G., and del Castillo, E. (2009) A Bayesian reliability approach to multiple response optimization with seemingly unrelated regression models *Qual. Tech. Quant. Management* **6**:353-369.
- Peterson, J.J. (2008) A Bayesian approach to the ICH Q8 definition of design space. *J. Biopharmaceutical Statistics* **18**: 959-975.
- Sambucini, V., and Piccanato, L., (2008) Likelihood and Bayesian approaches to inference for the stationary point of a quadratic response surface. *Canadian J.*