A Design Space to guarantee the long-term stability of a new formulation given production constraints: a Bayesian perspective

Pierre Lebrun – Bruno Boulanger, NCS 2012 Postdam

Context

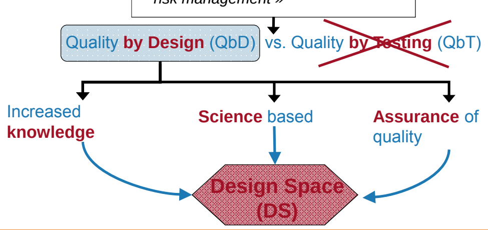
- Stability study
- Concentration (mg/mL) is evaluated with potency assays at t_0 and t_{lt} using 3 replicates
- Several classical stress conditions are assessed (S1, S2 and S3)
- The difference of concentrations $\delta_{0\text{-}lt}$ between t_0 and t_{lt} must be higher than -0.3
- Objective: Find *stable* formulation factor ranges out of 8 identified critical factors $X_1 \ldots X_8$

Quality by Design

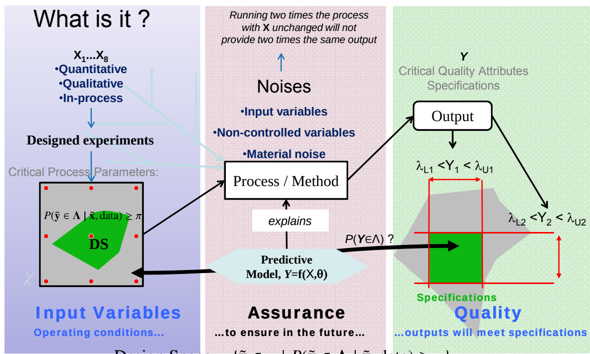
Quality by Design:

Regulatory Framework

« A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management »



Design Space

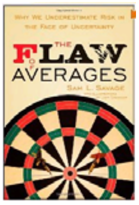


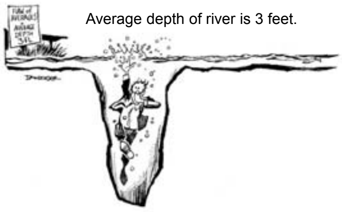
$$\text{Design Space} = \{\tilde{\mathbf{x}} \in \chi \mid P(\tilde{\mathbf{y}} \in \mathbf{\Lambda} \mid \tilde{\mathbf{x}}, \text{data}) \geq \pi\}$$

Focusing only on the mean (average)

Arlenda

(Courtesy of J.J. Peterson)





Average depth of river is 3 feet.

The Flaw of Averages:

Why We Underestimate Risk in the Face of Uncertainty

by Dr. Sam Savage

Predictive risk-based Design Space

Arlenda

*Running two times the process with **X** unchanged will not provide two times the same output*

The model used to explain the process is merely an approximation !

→ Generally, mean responses are used for optimization

✗ do not provide any clue about process reliability

✗ fail to give any information on how the process will perform in the future

✗ will certainly give disappointing and unexplained results for the future use of the method

✓ To solve these problems, one can use the posterior predictive distribution of the Critical Quality Attributes to express

- the guarantee of quality as a (posterior) probability of success
- the risk of non-quality...

QbD for formulation

Arlenda

- The process must provide, in its future use, **quality outputs**
  - e.g. during routine
- According to specifications derived from safety, efficacy, economical reasons
  - Whatever future conditions of use, that are not always perfectly controlled
  - Then, outputs should be **not sensitive** to minor changes
- This is **Quality by Design**
  - The way the process is developed leads to the product quality
  - This quality and the associated risks are assessed

Problem formalization

Arlenda

- Critical Quality Attributes
  - Difference of concentrations $\delta_{0\text{-}lt}$: Y
  - Format : a reportable result is the mean of three replicates
- Specifications
  - reportable results of $\delta_{0\text{-}lt} > -0.3$ mg/mL
- Factors
  - 8 formulations factors have been identified as Critical Process Parameters (CPP)
  - A Plackett-Burman design comprising 20 experiments has been conducted for every stress condition S1-S3

Predictive Bayesian Model

- Individual predictions will be drawn and the reportable results will be derived using simulations
- Take a lot of time to adjust your model
  - All your decisions are based upon it !
  - “Bad” model leads to very high predictive uncertainty
  - Take care not to overfit your model
- A multiple regression is adjusted
  - Will the attribute(s) be well explained by a (Normal) linear model ?
  - combination of variables, transformations ?
  - Y = **Xb** + e
- For this model, the posterior predictive distribution is identified in the Bayesian framework

Predictive density

- Model for multivariate regression $\underset{(n\times m)}{\mathbf{Y}} = \underset{(n\times p)}{\mathbf{X}}\ \underset{(p\times m)}{\mathbf{B}} + \underset{(n\times m)}{\mathbf{E}}$

$$\mathbf{y}_i \sim N_m(\mathbf{x}_i\mathbf{B},\boldsymbol{\Sigma}),\quad i = 1,\dots,n,$$

$$\mathcal{L}(\mathbf{B},\boldsymbol{\Sigma}\mid\mathbf{Y}) \propto |\boldsymbol{\Sigma}|^{\frac{-n}{2}}.\exp\left(-\frac{1}{2}tr\left[\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{XB})'(\mathbf{Y}-\mathbf{XB})\right]\right)$$

- Non-informative priors

$$p(\mathbf{B},\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)}$$

Informative priors

$$\mathbf{B}\mid\boldsymbol{\Sigma} \sim N_{p\times m}(\mathbf{B}_0,\boldsymbol{\Sigma},\boldsymbol{\Sigma}_0)$$
$$\boldsymbol{\Sigma} \sim W_1^{-1}(\boldsymbol{\Omega},\nu_0)$$

Posterior: $p(\mathbf{B},\boldsymbol{\Sigma}\mid\text{data}) \propto \mathcal{L}(\mathbf{B},\boldsymbol{\Sigma}\mid\mathbf{Y}).p(\mathbf{B}\mid\boldsymbol{\Sigma}).p(\boldsymbol{\Sigma})$

Predictive: $p(\tilde{\mathbf{y}}\mid\tilde{\mathbf{x}},\text{data}) = \int_{\boldsymbol{\Sigma}}\int_{\mathbf{B}} p(\tilde{\mathbf{y}}\mid\tilde{\mathbf{x}},\mathbf{B},\boldsymbol{\Sigma}).p(\mathbf{B},\boldsymbol{\Sigma}\mid\text{data}).d\mathbf{B}.d\boldsymbol{\Sigma}$

Non-informative:

$$\tilde{\mathbf{y}}\mid\tilde{\mathbf{x}},\text{data} \sim T_m\left(\tilde{\mathbf{x}}\hat{\mathbf{B}},\left(1+\tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X})^{-1}\tilde{\mathbf{x}}\right)\mathbf{A},\nu\right)$$

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$
$$\mathbf{A} = (\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})$$

Informative:

$$\tilde{\mathbf{y}}\mid\tilde{\mathbf{x}},\text{data} \sim T_m\left(\tilde{\mathbf{x}}\mathbf{M}_{\mathbf{B}\text{post}},\left(1+\tilde{\mathbf{x}}'(\mathbf{X}'\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}\tilde{\mathbf{x}}\right)(\boldsymbol{\Omega}+\mathbf{A}^*),\nu+n_0\right)$$

$$\mathbf{M}_{\mathbf{B}\text{post}} = \left(\mathbf{X}'\mathbf{X}+\boldsymbol{\Sigma}_0^{-1}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0\right)$$
$$\mathbf{A}^* = \mathbf{Y}'\mathbf{Y}+\mathbf{B}_0'\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0-(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)'(\mathbf{X}'\mathbf{X}+\boldsymbol{\Sigma}_0^{-1})^{-1}(\mathbf{X}'\mathbf{X}\hat{\mathbf{B}}+\boldsymbol{\Sigma}_0^{-1}\mathbf{B}_0)$$

When *m*=1, the multivariate *t* distribution simplifies to a univariate *t* distribution

Design Space computation

- One simulation for one factor setting
  - From the predictive distribution, sample 3 individual response predictions
  - Take the mean (reportable result) and compare to specification
- From many *n** simulations
  - Compute the MC estimate of the posterior probability of success

$$P(\mathbf{C}\tilde{\mathbf{Q}}\mathbf{A}\in\boldsymbol{\Lambda}\mid\tilde{\mathbf{x}},\text{data}) \simeq \frac{1}{n^*}\sum_{s=1}^{n^*} I(\mathbf{C}\tilde{\mathbf{Q}}\mathbf{A}^{(s)}\in\boldsymbol{\Lambda})$$

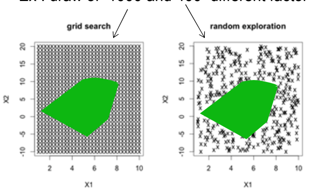
- For a grid over the factor setting
  - Draw maps of the posterior probabilities
  - Identify Design Space: $\{\tilde{\mathbf{x}}\in\chi \mid P(\tilde{\delta}_{0-lt} > -0.3\mid\tilde{\mathbf{x}},\text{data}) \geq \pi\}$

DoE considerations

- Unfortunately, not possible to explore every process parameter
  - DoE to analyze only the Critical Process Parameters
- Obviously, the analyst often believes that a lot of factors will impact his/her quality... and might be right about it !
- Computationally, there is a problem to represent high dimensional space of factors
  - Assume we want to explore a grid made from 10 points per factor...
  - ...10^8 conditions to explore !
- Parallelization, computer clusters, etc., are of no help in this case

Design Space representation

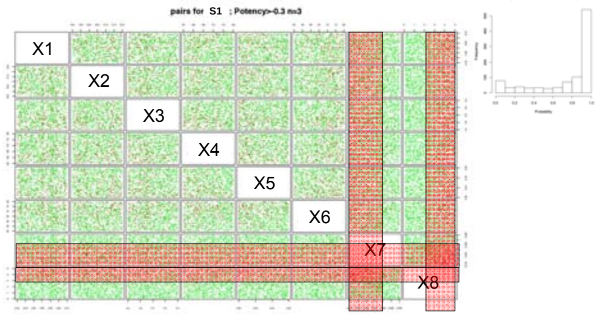
- A possibility is to explore the experimental domain by drawing randomly from a multivariate uniform distribution covering the space of factors
  - Ex : draw of 1000 and 400 different factor settings



- On each point, compute the (posterior) probability of success
- Then, create bivariate pair-plots of the factors

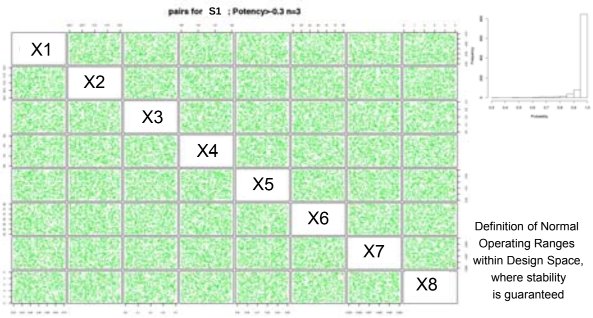
Results

- Pairs plot



Results

- Pairs plot



Conclusions

- Design Space is a tool build over DoE
  - The advantages of DoE are kept...
  - ...while fully taking into account all uncertainties to make sure the decision and the associated risks are controlled
- From the 8 process parameters, most of them were found “not so critical” and the risk-based optimization over only one of them allowed to improve and control the drug formulation to obtain satisfying stability given pre-defined specifications
- Pairs plots with random exploration of experimental domain can help when dimensionality is to high
  - If response(s) form is/are simple enough

Thank you !

pierre.lebrun@arlenda.com

www.arlenda.com