Challenges and Solutions in Drug Development

Optimization of a chromatographic separation by DOE approach

Dr. Lorenza Broccardo
QbD Senior Scientist

Bracco SpA, September 21th 2017
Contents

1. Introduction to DOE methodology
   • principal concepts
   • why is it useful?

2. application: optimization of a chromatography separation
   • problem formulation
   • the design
   • data analysis and results
   • what to do after optimization?

3. conclusions
   • comparing DOE/COST approach
   • benefit of DOE
Introduction to DOE methodology

A system study implies making experiments to acquire new information.

DOE (Design of Experiments) is a statistical methodology that provides workflows and instruments to:

1. Plan the experiments
2. Analyze the data
3. Find the solution
Experiments are planned according to a specific structure called “design”

Advantages adopting a so structured experiments’ organization:

• homogeneous sampling of the experimental domain
• evaluation of single factor effect and two factors interaction effect
• decisions are supported by a regression model with predictive power
Data are analyzed with the support of a regression model:

\[ y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + e \]

- coefficients like \( b_1 \) and \( b_2 \) provide information about single factor effect
- coefficients like \( b_{12} \) provide information about two factors interaction effect

Tools available for analyze the data and extract information:

- Coefficients plot
- N-Probability plot
- Replicates
- Histogram
- Interaction Plot

Many others!
Response Surface Model supports the decision making step

- the developed model is plotted to generate a surface that represents how Y variable relates with \( X_1 \) and \( X_2 \) variables

- which experimental conditions provide \( Y \) max?
- which experimental conditions provide \( Y \) min?
- which experimental conditions provide the most robust \( Y \)?
Response Surface Model supports the decision making step

Moreover: is there a better solution outside the tested experimental domain?

better solution could be find at lower level of variable $X_1$ and higher level of variable $X_2$. 

```latex
\text{find the solution}
```
DOE benefit summary

DOE is a more efficient methodology compared with the classical approach because:

• it drives the experimenter to set up an homogeneous sampling of the experimental domain

• it allows the evaluation of single factor effect and two factors interaction effect

• decisions are supported by a regression model (RSM)

Better system understanding is obtained by means of fewer experiments
Application

Optimization of a chromatography separation (HPLC) of a Hoodia Gordonii extract

Introduction

• This study was performed by University of Padova (Dott.ssa Elena Arnoldi, Dott. Stefano Dall’Acqua, Prof.ssa Gabbriella Innocenti) with the support of S-IN Soluzioni Informatiche

• Hoodia Gordonii is a plant growing naturally in South Africa and Namibia

• It’s extract is reach in Saponines, a class of chemical compounds of interest as dietary supplements and nutraceuticals

• The Hoodia Gordonii extract includes 10 identified Saponines of interest and several other molecules

• the HPLC method in place to separate the compounds of interest should be optimized in order to properly isolate the 10 Saponines

• The optimization have been performed with both approach \[ \text{COST} \] \[ \text{DOE} \] comparison
Problem formulation

Objective: optimization of the HPLC separation of 10 Saponines

Factors definition

• Several parameters may influence the retention times and the shape of the chromatographic peaks

• already available information on the system were used to properly select the factors to be investigates:
  ▪ eluent : mix of $\text{H}_2\text{O}/\text{C}_2\text{H}_3\text{N}$ ;
  ▪ eluent gradient is an important factor to improve the separation
  ▪ $\text{H}_2\text{O} \% = $ decreasing function of time
  ▪ $t(\text{final}) \text{H}_2\text{O} \% = 0$

• On the bases of the available these information, and with the intention to keep the number of experiments to perform as low as possible, we decided to
  ▪ study only the eluent gradient (we set $X = \text{H}_2\text{O} \%$ )
  ▪ describe $X$ as a parametric function of time ($X = f(t)$ see next slide)
  ▪ define experimentally the best parameters’ set
Gradient function \( X = f(t) \) definition

- \( t_1 \): end of first gradient [5; 25] min
- \( t_2 \): end of second gradient [25; 60] min
- \( X_1 (\text{H}_2\text{O}\%) \): eluent composition [0; 70] %

We studied the effect (on the chromatographic separation) of \( X_1 \), \( t_1 \) and \( t_2 \) variation setting up the experiments in this way:

The study includes 3 factors studied by a full factorial design which requires 8+3 experiments
Chromatogram and the parametric gradient function

Retention Time (min)

Response

$0.00$

$t_1$

$t_2$
**Response definition**

- Response = Y = resolution (R)
- goal: achieve a good resolution for all 10 Saponines peaks

However:

the resolution of 2 peaks is evaluated using the resolution definition

\[ R_{\text{critico}} = \frac{|t_{RA} - t_{RB}|}{\left(W_A + W_B\right)^2} \]

What about the resolution of 10 peaks?
We defined a parameter $R_{\text{score}}$, which summarize, in one value, the resolution of all peaks of interest in this way:

- for each Saponine peak, we considered the previous and consecutive signal and calculated the corresponding resolution values $R_i$

- compared the measured value $R_i$ with references values selected by University of Padova researchers:

  $R_{\text{opt}} = \text{resolution value that assures a good two consecutive peaks separation}$

  $R_{\text{min}} = \text{resolution value corresponding to the lowest acceptable separation between two consecutive peaks}$
then, we assigned to each experimental resolution \((R_i)\) a score according to the following rules:

\[
R_i \geq R_{\text{opt}} \quad \Rightarrow S_i = 1 \\
R_i < R_{\text{min}} \quad \Rightarrow S_i = 0 \\
R_{\text{min}} \leq R_i < R_{\text{opt}} \quad \Rightarrow S_i = \frac{R_i - R_{\text{min}}}{R_{\text{opt}} - R_{\text{min}}} \]

and calculated the “overall” resolution \((R_{\text{score}})\) according to the function:

\[
R_{\text{score}} = \frac{1}{N-1} \sum_{i=1}^{N} R_{si} 
\]

\(N = \) number of resolutions measured \\
\(R_{si} = \) score value

\(R_{\text{score}}\) is a number that summarizes the resolution of 10 signals of interest
Problem formulation summary:

Objective: HPLC optimization

Factors and corresponding variability intervals

- **t1 (time):** end of first gradient  [5; 25] min
- **t2 (time):** end of second gradient  [25; 60] min
- **X1 (H₂O%):** eluent composition at  \( t = t1 \)  [0; 70] %

Response:  \( R_{score} \)

Design: Full Factorial (8 + 3 experiments)

As a result of the problem formulation the worksheet (the list of the experimental condition to test) is obtained

![Worksheet Image]
Experimental session

11 experiments were performed and the corresponding value of $R_{score}$ were calculated

<table>
<thead>
<tr>
<th>Exp No</th>
<th>Exp Name</th>
<th>Run Order</th>
<th>Incl/Excl</th>
<th>H2O%</th>
<th>t1</th>
<th>t2</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N1</td>
<td>4</td>
<td>Incl</td>
<td>0</td>
<td>5</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>N2</td>
<td>2</td>
<td>Incl</td>
<td>70</td>
<td>5</td>
<td>25</td>
<td>0.769404</td>
</tr>
<tr>
<td>3</td>
<td>N3</td>
<td>3</td>
<td>Incl</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>0.755453</td>
</tr>
<tr>
<td>4</td>
<td>N4</td>
<td>11</td>
<td>Incl</td>
<td>70</td>
<td>25</td>
<td>25</td>
<td>0.897044</td>
</tr>
<tr>
<td>5</td>
<td>N5</td>
<td>8</td>
<td>Incl</td>
<td>0</td>
<td>5</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>N6</td>
<td>1</td>
<td>Incl</td>
<td>70</td>
<td>5</td>
<td>60</td>
<td>0.78946</td>
</tr>
<tr>
<td>7</td>
<td>N7</td>
<td>5</td>
<td>Incl</td>
<td>0</td>
<td>25</td>
<td>60</td>
<td>0.612196</td>
</tr>
<tr>
<td>8</td>
<td>N8</td>
<td>9</td>
<td>Incl</td>
<td>70</td>
<td>25</td>
<td>60</td>
<td>0.925271</td>
</tr>
<tr>
<td>9</td>
<td>N9</td>
<td>7</td>
<td>Incl</td>
<td>35</td>
<td>15</td>
<td>42.5</td>
<td>0.685987</td>
</tr>
<tr>
<td>10</td>
<td>N10</td>
<td>10</td>
<td>Incl</td>
<td>35</td>
<td>15</td>
<td>42.5</td>
<td>0.751919</td>
</tr>
<tr>
<td>11</td>
<td>N11</td>
<td>6</td>
<td>Incl</td>
<td>35</td>
<td>15</td>
<td>42.5</td>
<td>0.712842</td>
</tr>
</tbody>
</table>

Data analysis

Data were analyzed with the support of the software MODDE (Sartorius Stedim Data Analytics)
Clicking on the model, the data are fitted and the Summary of fit plot and the Coefficients plot are prompted.

**Summary of fit plot**
Summarize model’s parameters useful to evaluate:
- the model fitting capabilities ($R^2$)
- the model prediction power ($Q^2$)

**Coefficients plot**
The coefficients plot reports, as histograms, the model coefficients and supports us in understanding:
- single factor effects
- identify and understanding two factors interaction effects
Coefficients plot shows that

$H_2O\%$ and $t_1$ are the most impacting factors; increasing their level, the Resolution ($R_{\text{score}}$) increases.

Factor $t_2$ has no effect on Resolution.

The two factors interaction $H_2O\%*t_1$ and $t_1*t_2$ do not affect the response.

The two factors interaction $H_2O\%*t_1$ has a negative effect on Resolution (that is, increasing the level of both $H_2O\%$ and $t_1$, the additional negative contribution of $H_2O\%*t_1$ coefficient must be considered).
Model refinement

Eliminating the non significant coefficients a better model is obtained (higher $R^2$ and $Q^2$)

This is the best model to describe the system
Prediction

The Response Surface Model (the model plot) shows, graphically the relation between factors \((H_2O\%, t_1)\) and the response (Resolution: \(R_{score}\))

A detailed description of the response variation within all explored experimental domain is available

The best solution is not a point, but an area
Optimization of chromatographic separation conclusion - 1

**COST**
- area 1 was identified

**DOE**
- detailed description of $R_{\text{score}}$ within all tested experimental domain was obtained
- area 1 was confirmed
- area 1 is more extended (orange band)
- a better solution was identified (red area)
Optimization of chromatographic separation conclusion - 2

COST
- A feeling of the influence of each single factor on the response was obtained

DOE
- A numerical value of the effect of
  - each single factor
  - each two factors interaction
  was obtained
Optimization of chromatographic separation conclusion - 3

**COST**
- Useful but limited information were obtained by means of
  - more than 30 experiments
  - 2 months of working days

**DOE**
- Detained and completed information were obtained by means of
  - 11 experiments
  - 1 week of working days
What to do after optimization?

1. Explore a even better area outside the tested experimental domain using the acquired information (gradient techniques)

2. Select the area of interest inside the tested experimental domain and acquire more information about it adding a few selected experiments

3. Select the area of interest inside the tested experimental domain and perform a robustness testing (evaluation of response stability to small factors variation)
Conclusions

DOE (Design of Experiments) is a statistical methodology that provides a workflow to plan a set of high efficiency experiments and to analyze the data.

Data analysis is supported by a dedicated software which simplifies the procedures.

More useful information are acquired by less experiments

save time and money
Challenges and Solutions in Drug Development

thank you for your attention