The Use of Multiple Linear Regression to Find Acceptable Ranges for Manufacturing Materials

Mike Aylott
GlaxoSmithKline

26th September 2012

Introduction

- In recent years regulators have become increasingly keen for input manufacturing materials, as well as for the end-product tests, to have acceptance criteria
- Previously any correlations between input material and end-product quality attributes were simply accounted for by a line of best-fit
- For inhaled products, particles of exactly the right size must be delivered into the lungs
- Therefore the active pharmaceutical ingredient (API) particle size must be manufactured within a very tight specification.

The Analysis

Input Variables
- Particle Size Distribution (PSD)
  o (X10, X50, X90, Span)
- Specific Surface Area (SSA)

Output Variables
- Coarse Particle Mass (CPMass)
- Fine Particle Mass (FPMass)
- Very Fine Particle Mass (vFPMass)

The Objective

- "To construct a range for the input materials, that would maximise the chances of making a product with acceptable quality"
- These ranges are sometimes known as Proven Acceptable Ranges (PARs)
- If a PAR is too narrow, we may struggle to find suitable input material to make a product
- If a PAR is too wide, a large proportion of product batches would be rejected.

The Analysis

- How can the Input and Output parameters be related?
- Remember the Input parameters are **attributes**, not process parameters, so cannot be adjusted orthogonally
- There is inevitably collinearity between the PSD parameters: e.g. X50 cannot be larger than X90!

Possible Analysis Options

- A series of simple linear regression analyses
- Ellipse plots
- Multivariate analysis
- Forward selection / backwards elimination
- Ridge regression
- Multiple linear regression.

Extra Issues

- Limited data – only a few batches available
- Strong evidence of a random batch effect (there are multiple batches and multiple measurements within each batch)
- Some products can involve more than one API, so a consistent approach is needed within the product as well as across products.

PAR Setting Strategy

- A **flowchart** was devised to define PARs and ensure consistency across APIs/products
- The **mixed linear model** was used, dealing with the random batch effect
- Scientific judgement was used alongside statistical evidence
- If an input parameter was not selected in the model its PARs were based on process capability.

Model Selection

- One output variable was analysed at a time
- All possible permutations of the input parameters were selected as potential models
- These models were classified according to the "Akaike's Information Criterion" (AIC) statistic
  - This provides the "goodness of fit" statistic, while penalising for the number of terms present, to reward a parsimonious model
  - To compare AIC between mixed models with different fixed effects, the Maximum Likelihood method was used.

Mixed Linear Model

- This approach generated a prediction interval, providing a multidimensional acceptance region, thus maximising the chances of the product being of acceptable quality to the patient, and ensuring manufacturing quality is maintained
- This analysis is based on a default **85% level of confidence**, rather than the traditional 95%
  - This is considered to be a manufacturing risk, as the final drug product is still assessed for quality
  - In many cases this level could be increased.

Illustration of a One-Factor Model

- Here the Output parameter can be plotted against the Input parameter
- This provides a better understanding of the relationship and how the prediction intervals shape the PAR
- The PAR can be derived directly from the prediction interval.
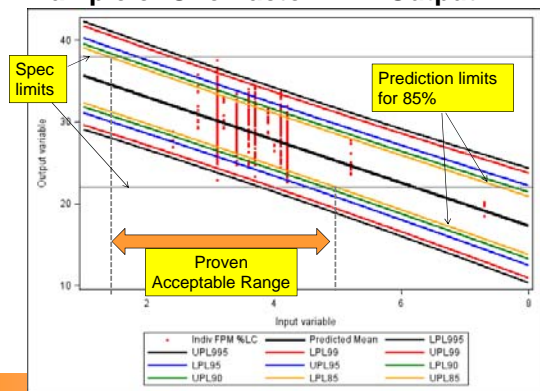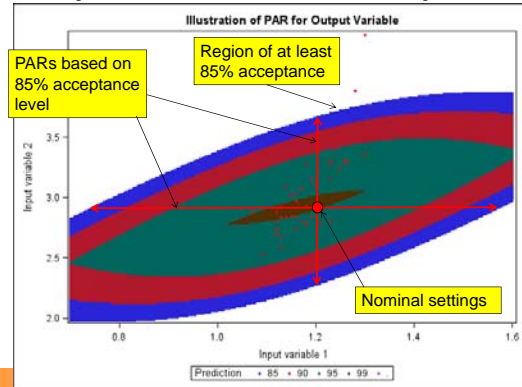
Example of One-Factor MLR Output

Illustration of a Two-Factor Model

- The Output parameter cannot be included in a two-dimensional plot
- Instead, the two Input parameters are plotted in a contour plot, while including:-
  - The input data
  - the % confidence of achieving the specification limits
  - The nominal setting of the process, to achieve the PAR
- Definition of a PAR in this situation – the range over which you can change one input attribute, whilst all others remain fixed.

Example of Two-Factor MLR Output

Reflection on this Approach MLR

- The approach enabled a consistent, rigorous analysis of the data
- Easy to display and understand graphically
- Concerns remain with its sensitivity, both over choice of model and changes to the dataset
- This approach could easily be modified to accommodate alternative definition of PARs
- There will inevitably be issues with analysing parameters that are so closely correlated.

Acknowledgements

- Mike Denham, GSK