africa

application notes

syrris

Note Number 9
Africa and Design of Experiments

Africa and Design of Experiments

1 Summary

Reaction optimisation work using traditional techniques can be complex, time consuming and expensive, but is a vital part of the development of all chemical processes. One approach is the use of statistical experimental design (DoE). This application note gives some guidance on DoE and it's use with the Africa system.

2 The Problem

In chemical processes, many factors will affect the overall performance of the reaction including temperature, time, stoichiometry, reagent choice, concentration etc. Classically, the chemist would look at the effect of these by varying one factor whilst maintaining the all others at a fixed levels, then choose the best level and repeat with the other factors (OFAT- One Factor At a Time).

With this approach, in the case of four factors at two levels each (low and high), this gives 2^4 or 16 possible experiments of which the chemist would do 8 – both levels of factor 1, then both level of factor 2 at the best level of factor 1 etc. If we suspect the factor effects are not linear, we would need to use three levels (low, middle and high) and this gives 3^4 or 81 experiments of which the chemist would do 12.

However, this approach is dangerously flawed:

- It assumes that all the factors are all independent, i.e. that the best level of one factor does not depend on the levels of the other factors. However, it is well know that e.g. time and temperature are often closely linked.
- It can leave large areas of the total experimental space unexplored. If we decide in the first two experiments that high temperature is best, we will gain no further information about low temperature during the remaining experiments.
- It is inefficient in terms of the number of experiments done to obtain the effects of the factors.

In principle, if we want the effects of four factors we are looking for the coefficients ($C_1 \ldots C_4$) of an equation of the form

$$\text{Response} = \text{Const} + C_1 \cdot \text{Fact}_1 + C_2 \cdot \text{Fact}_2 + C_3 \cdot \text{Fact}_3 + C_4 \cdot \text{Fact}_4$$

and this requires a minimum of five experiments. Unfortunately, not just any five from the 16 possible will give the right result - and there are 4368 possible choices of 5 from 16!

The difficulty in selecting an efficient set of experimental trials grows almost exponentially as the number and levels of factors grow and it is a problem that is not generally soluble by hand.

3 The Solution

The solution is DoE. This normally involves using a computer based statistical package to design your experimental trials and subsequently analyse your experimental data.

It is beyond the scope of this application note to give a comprehensive listing of all the methods available for DoE, but the information below gives a flavour for some of the approaches, along with some comments and references.

All but the simplest DoE work will require a design or statistical package to produce the design and a statistical package (or at least Excel) to analyse the results. There are also published tables of standard designs that are useful in some instances.

4 Types of Designs

There are many different types of designs that are more or less efficient in particular circumstances. Choice will depend on the objectives of the trail, the available statistical package, size of design (number of factors and levels), availability of previous experimental data and ease/cost of experiments. Below is a short list of some design types that are appropriate for chemical optimisation with Africa.

4.1 I-, A- and D-Optimal Designs

These designs generally represent the easiest and most efficient designs when larger numbers of factors are considered (3 or more). They have many advantages over classical factorial designs:

- Flexibility. You can set the size of a trial, and pick particular experiments that must/must not be included. You can omit 'impossible' factor combinations (eg. high temperature/low pressure etc.)
- You can re-use existing data. You can expand or in-fill an initial screening run to cover extended areas or new factors.
- You can produce very efficient designs to rapidly screen large experimental areas but with fewer limitations than Plackett-Burman designs.
- Optimal designs are particularly useful when resources are limited or there are constraints on factor settings.

Optimal designs are the designs of choice in most circumstances, but are always generated by a computer package and cannot normally be analysed 'by hand'. The design criteria are some mathematical measure of optimality in the underlying design matrix. I-optimality is theoretically most appropriate for this type of work, though historically D-optimality is most common.

4.2 Plackett-Burman Designs

Plackett-Burman (PB) designs are extremely efficient. The PB design in 12 runs, for example, may be used for an experiment containing up to 11 factors. However, they are normally used for screening experiments only because, in a PB design, main effects of factors cannot be separated from the two-factor interactions. These designs are useful for efficiently spotting large main effects, assuming all interactions are small when compared with the significant main effects but can be misleading and should be used with care.

4.3 Full and Fractional Factorials

These are classical designs that do all (full) or a fraction (½, ¼ etc. - fractional) of all the possible experiments. They are quick and easy to design and analyse by hand with small numbers of factors (2 or 3) but become less efficient with more factors, requiring more experiments than optimally designed methods.

4.4 Central Composite Designs

Central composite designs (CCD) designs start with a 2 level factorial or fractional factorial design and add "star" and centre points to estimate curvature. This can be a useful way to extend a preliminary 2 level screening design into a 3 or 5 level design. These are not always the most efficient designs but do have some replication built in, so final models are normally very reliable. Box-Behnken designs are similar, normally using slightly fewer trials. Again, these designs become less efficient when larger numbers of factors are examined.

5 Africa and DoE

Africa lends itself well to all the forms of DoE above. Reaction times, temperatures, stoichiometries and concentrations can all be automatically adjusted to allow multiple experiments to be performed automatically without user intervention. In particular, as temperatures are not limited by boiling point in Africa, a wider range can be explored than in standard laboratory equipment.

6 Useful References

- NIST Engineering Statistics Handbook
  Background information on many kinds of experimental design and analysis.
  http://www.itl.nist.gov/div898/handbook/pri/section5/pri521.htm
- MODDE
  Umetrics produce MODDE for DoE and analysis packages. They also provide training courses in experimental design and statistics.
  http://www.umetrics.com/software_modde.asp#About
- Design-Expert 6
  Stat-Ease produces two design and analysis package – Design Expert and Design Ease. The latter doesn't include D- and I- optimal methods.
  http://www.statease.com/dx6broch.html
- Statistica
  Statsoft produce a DoE add-on for Statistica that handles many designs including optimal deigns.
  http://www.statsoft.com/products/doe.html
- WebDOE
  This free DoE site can produce and analyse many kinds of design including I- and D-optimal designs. This site also has some useful background information on DoE.
  http://www.webdoe.cc/designs/
- JMP
  SAS Institute's JMP and ADX Interface provide both designs and analysis.
  http://www.jmp.com/index.shtml
- Minitab
  Minitab provides designs and analysis in a single package.
  http://www.minitab.com/