

# Design of Experiments

21 July to 15 August 2008

Report from the Organisers:

R. A. Bailey, B. Bogacka, H. Großmann (all Queen Mary, London)  
and D. Woods (Southampton)

## Background

Both the theory and practice of experimental design have fragmented over the last fifty years. Some of the divisions that have arisen are as follows.

Complex treatment models, such as factorial structures and response curves or surfaces for continuous treatments, typically go with a very simple structure on the experimental units. On the other hand, a simple additive model for discrete treatments is often twinned with a complex structure on the experimental units, either multi-stratum, or including correlations in space or time.

There may be one univariate observation per experimental unit, a multivariate observation, repeated measurements, or continuous time measurements.

The observational units may or may not be the same as the experimental units.

The response variable might be binary, categorical, counts, or continuous.

The model on the experimental units may be developed from knowledge of the application, or it may be based on randomization.

Designs may be constructed using search algorithms or combinatorial methods.

Designs can be sought which are locally optimum for particular values of the parameters, or which are optimal integrated over prior distributions of the parameters.

There may be various statistical properties of an experimental design which are optimized, such as parameter estimation, prediction of the response, hypothesis testing, or discrimination between competing models.

The design of experiments has been so successful in many different areas of application that there are now different traditions in different areas, such as biology, agriculture, engineering, psychology, clinical trials. For example, the recent rapid development of microarray experiments in genomics not only poses its own peculiar problems but has often ignored the knowledge base in design of experiments. The organisers saw an urgent need for cross-fertilization between the different areas.

The aim of the programme was to synthesize different approaches to the design of experiments, so that people in each area could incorporate what is best from other areas. It focussed on three methodological topics and three application areas. We chose topics where we anticipated that progress could be made in a short time, bringing together different approaches to tackle some specific problems, both applied and theoretical.

The methodological topics were multi-stratum experiments, multi-tiered experiments, and the design of experiments for non-linear models. Each is already used in several areas of application, but often with different vocabulary. The idea was to get together statisticians from the different areas so that they could pool expertise. Because the underlying methodology is the same, we believed that it would be possible to learn from each other, incorporate the best from each area and produce a more unified theory.

Our chosen application areas were genomics, computer experiments and clinical trials. Each of these is currently lively, and has interesting problems to be brought to the attention of a wider audience of design researchers. We hoped to make substantial progress simply by getting the application people together with the design people. Not only would specialist design knowledge be used to improve the design of experiments in the application area; particular expertise from the application area might turn out to be useful in the design of experiments in other areas.

## Programme Structure

The programme opened with a workshop on *Advanced Topics in Design of Experiments*, 21–25 July 2008. This had three goals. The first was to introduce researchers in different parts of the design of experiments to parts of the subject outside their own speciality. Three short courses were given: *Multi-stratum experiments*, *Optimal design for linear and non-linear models* and *Multi-tiered experiments*.

The second goal was to introduce three specific application areas where good design is needed. A theme day was devoted to each of *Experiments in genomics and proteomics*, *Computer experiments* and *Clinical trials*, with both statisticians and scientists explaining the topic.

The third goal was to introduce students and other interested people to the subject. This workshop was attended by 29 invited participants and 10 others.

The middle two weeks had no timetabled activities. The purpose was for the participants to work intensively together, to develop what they had learnt from the first week as well as to continue existing research projects. There were 26 invited participants during this phase.



R. A. Fisher, one of the pioneers of design of experiments, was a Fellow of Caius College. A. F. W. Edwards kindly entertained participants at Caius one evening, showing them the portrait of Fisher and the stained glass window of the Latin square on the cover of Fisher's book *The Design of Experiments*.

The final week of the programme was devoted to the workshop *Designed Experiments: Recent Advances in Methods and Applications (DEMA2008)*, 11–15 August 2008. This was the successor to the DEMA conference held in Southampton in 2006. There were plenary talks on randomisation and on design of two-phase experiments; theme days on genomics and proteomics, clinical trials, and computer experiments; and sessions on multi-stratum experiments, design for correlated data, block designs, design for non-linear models, and design construction and optimality. There were also two poster sessions, each preceded by a much-appreciated “poster storm”. This was attended by 41 invited participants and 64 others. Participants came from the pharmaceutical industry, software houses, engineering companies, an oceanography centre and medical research establishments, as well as mathematics and statistics departments.

## Outcomes and Achievements

The short courses in the first week were pitched at just the right level for the audience to engage with the material: this was as true for experts from other specialities as it was for PhD students. Awkward questions from people in different areas provoked useful discussion right from the start. The input from the three application areas enabled participants “to focus on the right design problems” (Bogacka). The introducing statisticians from these areas attended throughout, and “found the whole week extremely valuable” (Speed).

Many conversations in the middle two weeks were sparked off by this introductory material, and it was clear that participants were seeking information from each other as well as from “the excellent library” (Kunert). Several multi-hour round-table discussions were held with a dozen participants, of whom half shared expertise while the other half learnt. Small tutorials seemed to be in progress all the time.

Morgan's comment is typical: “I have made definitive progress on several problems, learned the fundamentals of several design areas in which my knowledge was sorely lacking, and laid the groundwork for future working relationships.” Some of the research achieved during the programme was already under way, and simply needed the calm and supportive environment of the INI for its completion. Most participants learnt about new areas, and started new work. The talks by Bailey, Bogacka, Challenor, Gilmour, Kunert and Morgan (at least) in DEMA2008 were based on work done during the programme.

Most exciting was the number of new collaborations. Joint work and discussions included the following. A. C. Atkinson, B. Bogacka and D. Uciniski: combining ethics with efficient estimation in Phase I clinical trials. A. C. Atkinson, B. Bogacka and M. Patan: design of experiments for non-linear models. R. A. Bailey, C. J. Brien and D. Woods: experiments for human–computer interaction. R. A. Bailey, C. J. Brien, C.-S. Cheng, R. Mee and P.-W. Tsai: experiments for multi-stage batch reprocessing. R. A. Bailey, P. J. Cameron, C.-S. Cheng and J. P. Morgan: the combinatorics of optimal design. R. A. Bailey, C.-S. Cheng, A. M. Dean, S. G. Gilmour, P. Goos, J. P. Morgan and

L. A. T. Trinca: the equivalent estimation property of some multi-stratum response surface designs. S. G. M. Biedermann, H. Dette and D. Woods: optimal designs for multivariate spline models. B. Bogacka, S. G. Gilmour, M. Latif and L. A. T. Trinca: design for enzyme kinetic models. C. J. Brien, A. Lynch and T. P. Speed: two-phase experiments in genomics. P. G. Challenor and D. Woods: sequential design for Gaussian process models. P. G. Challenor, A. M. Dean, H. Maruri-Aguilar, L. M. Moore and H. P. Wynn: design of computer experiments. C.-S.-Cheng and J. P. Morgan: incomplete-block designs. C.-S.-Cheng and L. M. Moore: nested factorial designs. A. M. Dean and L. M. Moore: trend-free factorial designs. A. M. Dean and J. P. Morgan: robust designs for marketing. A. M. Dean, H. Maruri-Aguilar and H. P. Wynn: supersaturated algebraic models. H. Dette and J. Kunert: optimal design for correlated models. H. Dette and H. Maruri-Aguilar: optimal design for hierarchical models. H. Dette and R. Schwabe: optimal design for non-linear models with random effects. S. G. Gilmour and P. Goos: Bayesian analysis of multi-stratum response surface designs. S. G. Gilmour, P. Goos and H. Großmann: randomization-based analysis of multi-stratum response surface designs. S. G. Gilmour, P. Goos and J. P. Morgan: factors which are randomized but not reset. S. G. Gilmour, H. Großmann and R. Schwabe: discrete choice experiments. S. G. Gilmour, L. A. T. Trinca and P.-W. Tsai: model-robust criteria for response surface studies. P. Goos and H. Großmann: order effects in paired comparison experiments. B. Jones and D. Woods: graphical assessment of designs for non-linear models. R. W. Payne, P. van de Ven and D. Woods: designs for hierarchical generalized linear models. L. Pronzato and H. P. Wynn: sequential design of clinical trials. E. D. Schoen and P. van de Ven: Bayesian optimal experiments. R. Schwabe and H. P. Wynn: Fourier regression designs. P. van de Ven and D. Woods: block designs for non-normal data.

As Gilmour commented: “I am leaving the INI with considerably more unfinished work than I arrived with. I regard this as a sign of a very successful programme.”

DEMA2008 was “one of the most successful conferences on design of experiments that I had ever attended” (Cheng) and “probably the best organised conference I have attended” (Godolphin). Much useful interaction took place between the workshop participants and the long-term participants. The geophysics group with A. Curtis made contact with mainstream design of experiments. K. Baggerly, R. A. Bailey, S. G. Gilmour, K. Kerr, V. Lima-Passos, A. Lynch and B. Parker discussed the design of experiments in genomics, and have already collaborated on two experiments. As the result of the talks they gave, R. A. Bailey, T. P. Speed and C. Vivacqua are advising the US National Institutes of Health on a large experiment on the collection and storage of blood. S. G. Gilmour spoke to M. Vandebroek about her work on choice experiments; S. G. Biedermann had useful discussions with M. Latif, B. Maus, K. Roth, M. Stehlik and P. van de Ven; as did A. C. Atkinson with H. Dette, A. Giovagnoli, S. Leonov and D. Woods. Another outcome is that A. C. Atkinson intends to organize a session at the Joint Statistical Meetings in Washington DC with D. Woods as an invited speaker.

It is likely that most of the research continued or initiated during the programme will lead to publications: ask us again in one year’s time!