Open access • Journal Article • DOI:10.1080/08982112.2018.1546398

Discussion on “Søren Bisgaard's contributions to Quality Engineering: Design of experiments” — Source link

Peter Goos, Peter Goos, Eric D. Schoen

Institutions: Katholieke Universiteit Leuven, University of Antwerp

Published on: 12 Apr 2019 - Quality Engineering (Taylor & Francis)

Discussion of “Soren Bisgaard’s Contributions to Quality Engineering: Design of Experiments”

Peter Goos*
Faculty of Bioscience Engineering and Leuven Statistics Research Centre (LSTAT), KU Leuven, Belgium
Faculty of Applied Economics and StatUa Center for Statistics, University of Antwerp, Belgium

Eric D. Schoen
Faculty of Bioscience Engineering, KU Leuven, Belgium
TNO, Zeist, Netherlands

November 5, 2018

1 Introduction

In his review of Soren Bisgaard’s contributions in the field of design of experiments, Vining (2018) devotes Section 4 to Bisgaard’s 2000 Journal of Quality article ‘The Design and Analysis of $2^{k-p} \times 2^{q-r}$ Split-Plot Experiments’. Throughout that section, Vining (2018) makes several statements concerning Bingham and Sitter’s work on fractional factorial split-plot designs, which appeared in various articles between 1999 and 2001 (Bingham and Sitter, 1999a,b, 2001), following an earlier publication by Huang et al. (1998) on the same topic. Vining (2018) also makes a few statements concerning the work of Letsinger et al. (1996) and the use of the optimal experimental design framework for creating split-plot designs. In this discussion, we argue that Vining misrepresents both the Bingham and Sitter approach and the optimal design framework for split-plot designs and we discuss some of the useful follow-up work that has been done, building on Bingham and Sitter (1999a,b, 2001) and Bisgaard (2000).

2 The work of Bingham and Sitter

Like Huang et al. (1998), Bingham and Sitter (1999a,b, 2001) construct two-level fractional factorial split-plot designs using the minimum aberration criterion, which is a refinement of the resolution criterion for selecting fractional factorial designs (Fries and Hunter, 1980). Two designs may have the same resolution, but perform differently in terms of the aberration criterion. In designs with a smaller aberration, roughly speaking, fewer low-order effects are aliased with other low-order effects.

In this discussion, we focus on the following statements of Vining (2018):

A Page 15: “Maximizing the overall resolution of the experiment, taken to its logical conclusion, is to have as many whole plots as possible with as few sub-plots as possible, making the resulting design look as close as possible to a completely randomized design, ...”

B Page 15: “The Bingham and Sitter approach uses a search algorithm to find the minimum aberration design over all of the estimable effects.”

*Address correspondence to Prof. dr. Peter Goos, KU Leuven, Faculty of Bioscience Engineering, Kasteelpark Arenberg 30, 3001 Heverlee, Belgium. E-mail: peter.goos@kuleuven.be.

C Page 16: “Soren makes the point that it is a mistake to expect the design resolution for the whole-plot effects to be as high as for the sub-plot factors. This latter point is especially important because the algorithmic approach (of Bingham and Sitter) attempted to make both the whole-plot and sub-plot resolution the same.”

D Page 16: “A general algorithmic approach, such as the one proposed by Bingham and Sitter, typically cannot take into account the specific research questions and the proper trade-offs to consider in creating the specific experimental plan.”

E Page 18: “The criterion used by the Bingham and Sitter work, along with the criteria in much of the optimal split-plot literature, suggests as a logical conclusion that one should use only one sub-plot per whole plot, which converts the split-plot experiment into a completely randomized design! Something is very amiss with such a conclusion.The problem is the failure to recognize that there is less information about the whole plots, in some cases dramatically less.”

Bingham and Sitter (1999a) present a generic construction of minimum aberration two-level fractional-factorial split-plot designs. More specifically, they completely enumerate all possible (non-isomorphic) fractional factorial split-plot designs. For each number of whole plots and sub-plots, and for each number of whole-plot and sub-plot factors, they then select and report the minimum aberration design. Therefore, their designs guarantee that as few low-order effects are aliased with other low-order effects as possible.

Bingham and Sitter (1999a) basically first fix the number of whole plots and sub-plots per whole plot. Second, they set up the whole-plot basic factors and the sub-plot basic factors. Third, they add the whole-plot added factors and the sub-plot added factors in all possible ways, subject to two restrictions:

1. Whole-plot added factors can only be constructed based on interactions involving whole-plot basic factors. Otherwise, the whole-plot added factors' levels would not remain constant within a whole plot, which would violate the split-plot structure.
2. Sub-plot added factors can be constructed based on interactions between whole-plot basic factors and at least one sub-plot basic factor or based on interactions involving sub-plot basic factors only.

As a result, in the work of Bingham and Sitter (1999a), the number of whole plots and sub-plot is an input parameter, as is the total number of observations. Also, the whole-plot factors and sub-plot factors are identified at the start of the design construction. Consequently, there is no recommendation whatsoever to have as many whole plots as possible, and the corresponding statements in citations (A) and (E) from Vining (2018) are wrong.

We now study Vining's statements (B) and (C). Suppose that Vining's statements (B) and (C) are true and that we face a design problem with 3 whole-plot factors, 3 sub-plot factors and 16 observational units, as in the example on page 19 of his review paper. Statement (B) would imply that the search algorithm would find the resolution-IV minimum aberration 2^{6-2} design in all cases. This is in contradiction with the actual outcomes of the search algorithm, which are reported in Table 2 of Bingham and Sitter (1999a). In that table, two split-plot design options are presented. The first one, designated design 3.3.0.2, involves 8 whole plots of 2 sub-plots each. The second one, designated design 3.3.1.1, involves 4 whole plots of 4 sub-plots each. Design 3.3.0.2 indeed corresponds with the minimum aberration 2^{6-2} design of resolution IV. Design 3.3.1.1, however, in no way corresponds to the minimum aberration 2^{6-2} design. Instead, it corresponds exactly to the design presented on page 21 of Vinings paper. This illustrates that the algorithm of Bingham and Sitter (1999a) finds minimum aberration designs subject to the constraints on the numbers of whole plots, sub-plots, whole-plot factors and sub-plot factors imposed by the experimenter. In addition, the whole-plot and sub-plot resolutions are not the same. We conclude that Vining's statements (B) and (C) do not represent a realistic view of the work of Bingham and Sitter.

By using the minimum aberration criterion without distinguishing between main effects and interactions of whole-plot plot factors, main effects ant interactions of sub-plot factors and interactions between whole-plot and sub-plot factors, Bingham and Sitter (1999a,b, 2001) assume that, to the experimenter, all effects are of equal interest. This is typical for screening experiments in which there is no prior information concerning the importance of the factors or the effects, and in which there are no specific effects on which the experimenter wants to focus. Obviously, in the event some specific effects are of more interest to the experimenter than others, the design selection criterion has to be adjusted to reflect this. Bingham and Sitter (2001) certainly realize this, writing that "...the [minimum aberration] criterion treats all factors and effects of the same order equally. If one has particular interest in a subset of effects, then some other criterion is required." This complements their earlier statement that "there are cases in which it is preferable to use the second or third best [fractional factorial split-plot design]" (Bingham and Sitter, 1999a). Thus, we believe that Vining's statement (D) is unjustified.

Our view on the work of Bingham and Sitter (1999a,b, 2001), and later Bingham and Sitter (2003) and Bingham et al. (2004), is that they present extremely useful catalogs of split-plot designs, which allow practitioners to look up sensible designs without having to do the algebra with design generators themselves. Unless the practitioners have a very specific interest in certain factors or effects, the designs offered in the catalogs will be good starting points for designing their split-plot screening experiments. For the vast majority of the practitioners, we believe that the published catalogs are even essential to be able implement the approach of Bisgaard (2000). To illustrate this, suppose that a practitioners needs to design a split-plot experiments involving 2 whole-plot factors, 8 sub-plot factors, 4 whole plots and 4 sub-plots per whole plot. For this complex a problem, it is extremely hard for a practioner to turn the general principles presented by Bisgaard (2000) into a useful design. If he/she does succeed in making one design, it is likely that the practitioner would be satisfied and would not search for any possibly better design. We therefore consider it very useful that Bingham and Sitter, in their Technometrics paper (Bingham and Sitter, 1999a), indicate that there are three different minimum aberration designs for this case. It is even more useful for practitioners to consult Bingham and Sitter's Journal of Quality Technology paper (Bingham and Sitter, 2001), because, there, they also discuss the differences between the three designs in detail. In particular, they explain that the designs differ in the number of sub-plot factor interactions that are confounded with the whole plots. The three designs have 7, 8 and 12 sub-plot interactions that are confounded with whole plots. More specifically, 4, 4 and 8 of these interactions are aliased with main effects of the whole-plot factors. We can therefore question whether any of the three designs is likely to give useful results. However, given the restrictions in whole plots, sub-plots, whole-plot factors and sub-plot factors, these three designs are the best we can get, and Bingham and Sitter (2001) do an admirable job discussing the properties of the designs. This example with three different designs additionally shows that the statement in the last sentence of citation (E) is not justified. Indeed, the fact that there is less information about the whole plots, in some cases dramatically less, is clearly recognized here.

3 Optimal design approach for split-plot experiments

Vining (2018) states that the analysis approach of Letsinger et al. (1996) for data from split-plot experiments is "extremely flawed" (page 15), without providing any arguments for that strong statement. This is remarkable given that the analysis approach of Letsinger et al. (1996) is standard practice in many application areas of statistics. What Letsinger et al. (1996) recommend is nothing but generalized least squares for response surface models based on split-plot data and restricted maximum likelihood for estimating the variance components in split-plot models. This approach, which was mentioned already in an example in Littell et al. (1996), recommended by Gilmour and Trinca (2000) for blocked experiments and studied in detail by Langhans et al. (2005) and Goos et al. (2006), has become the standard approach to analyze data from split-plot response surface experiments and has been implemented in the best industrial statistics software packages. It generalizes the traditional analysis of variance, allows experimenters to go beyond traditional balanced, orthogonal experimental designs, and can be used to generate pure-error estimates and conduct lack-of-fit tests too (Goos and Gilmour, 2017).

Finally, like Bingham and Sitter's work, most papers on optimal design of split-plot experiments assume that the number of whole plots as well as the number of sub-plots per whole plot is dictated by the experimental situation (Goos and Vandebroek, 2003; Jones and Goos, 2007, 2012; Mylona et al., 2014; Sambo et al., 2014; Trinca and Gilmour, 2015, 2017; Borrotti et al., 2017). In addition, it has been shown, within the optimal experimental design framework, that optimal split-plot designs outperform completely randomized designs. For instance, Goos and Vandebroek (2004) used an algorithm that optimizes the number of whole plots (with an option to impose an upper bound on that number) as well as the number of sub-plots within a whole plot, and found that a completely randomized design is generally suboptimal. So, optimal experimental design does not at all suggest to use only one sub-plot per whole plot either.

4 Follow-up work

When performing a split-plot experiment in the context of robust parameter design, it is often natural to focus more on the whole-plot-by-sub-plot interaction effects because the whole-plot factors generally correspond to noise factors, the sub-plot factors correspond to control factors, and control-by-noise interactions are key in designing robust products and processes. To deal with such a situation, Bingham and Sitter (2003) presented an alternative version of the minimum aberration criterion for robust parameter experiments. In their nuanced discussion, Bingham and Sitter (2003) once more point out that there may be instances in which designs other than the minimum aberration designs would be sensible choices too.

Bingham and Sitter (1999a,b, 2001) assume that every combination of whole-plot factor levels is visited exactly once. Bingham et al. (2004) relax this assumption and allow the combinations of whole-plot factor levels to be visited more than once. They present an extra catalog of useful fractional factorial split-plot designs.

All of this work of Bingham and Sitter (1999a,b, 2001, 1999a) and Bingham et al. (2004) involves design generators and requires the number of experimental runs as well as the number of whole plots to be a power of 2. Sartono et al. (2015) present a general methodology to construct orthogonal two-level fractional factorial split-plot designs, based on linear programming. This methodology combines two different two-level orthogonal arrays, one for the whole-plot factors and one for the sub-plot factors, and can be used whenever the number of whole plots is a multiple of 4 and the number of runs within a whole plot is even.

In recent years, strip-plot designs have received quite a bit of attention as well. Vivacqua and Bisgaard (2004, 2009) constructed regular fractional factorial strip-plot designs, for which the number of runs, the number of rows and the number of columns are all powers of 2. These articles inspired Goos and Jones (2011) to devote a chapter to strip-plot experiments in their book on optimal experimental design. Arnouts et al. (2010, 2013) present an alternative construction based on optimal experimental design methodology. The combination of split-plot designs and blocking has been discussed in Capehart et al. (2012), while staggered-level designs have been proposed by Arnouts and Goos (2012, 2015) as an alternative to split-plot designs when there are multiple so-called hard-to-change factors.

Several sets of authors have also presented general construction methods for two-level multi-stratum designs, which is a class of designs that includes split-plot, split-split-plot and strip-plot designs, among others. These methods are described in Bingham et al. (2008) and Cheng and Tsai (2011).

All of these published paper show that the attention Soren Bisgaard helped drawing to split-plot designs has resulted in a substantial amount of useful follow-up research, to cover experimental scenarios other than those he studied.

About the authors

Prof. dr. Peter Goos is Full Professor at the Faculty of Bioscience Engineering and the Leuven Statistics Research Centre (LStat) of the KU Leuven and at the Faculty of Applied Economics and StatUa Center for Statistics of the University of Antwerp. He is a Senior Member of the American Society for Quality. His email address is peter.goos@biw.kuleuven.be.

Prof. dr. Eric Schoen is Senior Statistical Consultant at TNO and Guest Professor at the Faculty of Bioscience Engineering of the KU Leuven. His email address is eric.schoen@kuleuven.be.

References

Arnouts, H. and Goos, P. (2012). Staggered-level designs for experiments with more than one hard-to-change factor. *Technometrics*, 54(4):355–366.

Arnouts, H. and Goos, P. (2015). Staggered-level designs for response surface modeling. *Journal of Quality Technology*, 47(2):156–175.

Arnouts, H., Goos, P., and Jones, B. (2010). Design and analysis of industrial strip-plot experiments. *Quality and Reliability Engineering International*, 26(2):127–136.

Arnouts, H., Goos, P., and Jones, B. (2013). Three-stage industrial strip-plot experiments. *Journal of Quality Technology*, 45(1):1–17.

Bingham, D., Sitter, R., Kelly, E., Moore, L., and Olivas, J. D. (2008). Factorial designs with multiple levels of randomization. *Statistica Sinica*, 18:493–513.

Bingham, D. R., Schoen, E. D., and Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Journal of the Royal Statistical Society, Ser. C (Applied Statistics)*, 53:325–339. Corrigendum, 54, 955-958.

Bingham, D. R. and Sitter, R. R. (1999a). Minimum-aberration two-level fractional factorial split-plot designs. *Technometrics*, 41:62–70.

Bingham, D. R. and Sitter, R. R. (1999b). Some theoretical results for fractional factorial split-plot designs. *Annals of Statistics*, 27:1240–1255.

Bingham, D. R. and Sitter, R. R. (2001). Design issues in fractional factorial split-plot designs. *Journal of Quality Technology*, 33:2–15.

Bingham, D. R. and Sitter, R. R. (2003). Fractional factorial split-plot designs for robust parameter experiments. *Technometrics*, 45:80–89.

Bisgaard, S. (2000). The design and analysis of $2^{k-p} \times 2^{q-r}$ split plot experiments. *Journal of Quality Technology*, 32:39–56.

Borrotti, M., Sambo, F., Mylona, K., and Gilmour, S. (2017). A multi-objective coordinate-exchange two-phase local search algorithm for multi-stratum experiments. *Statistics and Computing*, 27(2):469–481.

Capehart, S. R., Keha, A., Kulahci, M., and Montgomery, D. C. (2012). Generating blocked fractional factorial split-plot designs using integer programming. *International Journal of Experimental Design and Process Optimisation*, 3(2):111–132.

Cheng, C.-S. and Tsai, P.-W. (2011). Multistratum fractional factorial designs. *Statistica Sinica*, 21:1001–1021.

Fries, A. and Hunter, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics*, 22:601–608.

Gilmour, S. G. and Trinca, L. A. (2000). Some practical advice on polynomial regression analysis from blocked response surface designs. *Communications in Statistics: Theory and Methods*, 29:2157–2180.

Goos, P. and Gilmour, S. G. (2017). Testing for lack of fit in blocked, split-plot, and other multi-stratum designs. *Journal of Quality Technology*, 49(4):320–336.

Goos, P. and Jones, B. (2011). *Design of Experiments: A Case Study Approach.* New York: Wiley.

Goos, P., Langhans, I., and Vandebroek, M. (2006). Practical inference from industrial split-plot designs. *Journal of Quality Technology*, 38:162–179.

Goos, P. and Vandebroek, M. (2003). D-optimal split-plot designs with given numbers and sizes of whole plots. *Technometrics*, 45:235–245.

Goos, P. and Vandebroek, M. (2004). Outperforming completely randomized designs. *Journal of Quality Technology*, 36:12–26.

Huang, P., Chen, D., and Voelkel, J. (1998). Minimum-aberration two-level split-plot designs. *Technometrics*, 40:314–326.

Jones, B. and Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs. *Journal of the Royal Statistical Society. Series C*, 56:347–364.

Jones, B. and Goos, P. (2012). I-optimal versus D-optimal split-plot response-surface designs. *Journal of Quality Technology*, 44:85–101.

Langhans, I., Goos, P., and Vandebroek, M. (2005). Identifying effects under a split-plot design structure. *Journal of Chemometrics*, 19:5–15.

Letsinger, J. D., Myers, R. H., and Lentner, M. (1996). Response surface methods for bi-randomization structures. *Journal of Quality Technology*, 28:381–397.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary: SAS Institute.

Mylona, K., Goos, P., and Jones, B. (2014). Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation. *Technometrics*, 56(2):132–144.

Sambo, F., Borrotti, M., and Mylona, K. (2014). A coordinate-exchange two-phase local search algorithm for the D- and I-optimal designs of split-plot experiments. *Computational Statistics & Data Analysis*, 71:1193 – 1207.

Sartono, B., Goos, P., and Schoen, E. (2015). Constructing general orthogonal fractional factorial split-plot designs. *Technometrics*, 57(4):488–502.

Trinca, L. A. and Gilmour, S. G. (2015). Improved split-plot and multistratum designs. *Technometrics*, 57(2):145–154.

Trinca, L. A. and Gilmour, S. G. (2017). Split-plot and multi-stratum designs for statistical inference. *Technometrics*, 59(4):446–457.

Vivacqua, C. A. and Bisgaard, S. (2004). Strip-block experiments for process improvement and robustness. *Quality Engineering*, 16:495–500.

Vivacqua, C. A. and Bisgaard, S. (2009). Post-fractionated strip-block designs. *Technometrics*, 51:47–55.