UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper M09/20
Methodology

A Comparison Of Design And Model Selection Methods For Supersaturated Experiments

Christopher J. Marley, David C. Woods

Abstract

Various design and model selection methods are available for supersaturated designs having more factors than runs but little research is available on their comparison and evaluation. In this paper, simulated experiments are used to evaluate the use of E(s2)-optimal and Bayesian D-optimal designs, and to compare three analysis strategies representing regression, shrinkage and a novel model-averaging procedure. Suggestions are made for choosing the values of the tuning constants for each approach. Findings include that (i) the preferred analysis is via shrinkage; (ii) designs with similar numbers of runs and factors can be e_ective for a considerable number of active e_ects of only moderate size; and (iii) unbalanced designs can perform well. Some comments are made on the performance of the design and analysis methods when e_ect sparsity does not hold.

A comparison of design and model selection methods for supersaturated experiments

Christopher J. Marley and David C. Woods[1]
Southampton Statistical Sciences Research Institute,
University of Southampton,
Southampton, S017 1BJ, UK

Various design and model selection methods are available for supersaturated designs having more factors than runs but little research is available on their comparison and evaluation. In this paper, simulated experiments are used to evaluate the use of $E(s^2)$-optimal and Bayesian D-optimal designs, and to compare three analysis strategies representing regression, shrinkage and a novel model-averaging procedure. Suggestions are made for choosing the values of the tuning constants for each approach. Findings include that (i) the preferred analysis is via shrinkage; (ii) designs with similar numbers of runs and factors can be effective for a considerable number of active effects of only moderate size; and (iii) unbalanced designs can perform well. Some comments are made on the performance of the design and analysis methods when effect sparsity does not hold.

Keywords: Bayesian D-optimal designs; $E(s^2)$-optimal designs; Effect sparsity; Gauss-Dantzig selector; Main effects; Screening; Simulation

1. Introduction

A screening experiment investigates a large number of factors to find those with a substantial effect on the response of interest, that is, the active factors. If a large experiment is infeasible, then using a supersaturated design in which the number of factors exceeds the number of runs may be considered. This paper investigates the performance of a variety of design and model selection methods for supersaturated experiments through simulation studies.

Supersaturated designs were first suggested by Box (1959) in the discussion of Satterthwaite (1959). Booth and Cox (1962) provided the first systematic construction method, and made the columns of the design matrix as near orthogonal as possible through the $E(s^2)$ design selection criterion

[1]Corresponding author. Tel.: +44 23 8059 5117; fax: +44 23 8059 5147.
Email: D.Woods@soton.ac.uk

(see Section 2.1.1). Interest in design construction was revived by Lin (1993) and Wu (1993), who developed methods based on Hadamard matrices. Recent theoretical results for $E(s^2)$-optimal and highly efficient designs include those of Nguyen and Cheng (2008). The most flexible design construction methods are algorithmic: Lin (1995), Nguyen (1996) and Li and Wu (1997) constructed efficient designs for the $E(s^2)$ criterion. More recently, Ryan and Bulutoglu (2007) provided a wide selection of designs that achieved lower bounds on $E(s^2)$, and Jones et al. (2008) constructed designs using Bayesian D-optimality. For a review of supersaturated designs, see Gilmour (2006).

The challenges in the analysis of data from supersaturated designs arise from correlations between columns of the model matrix and the fact that the main effects of all the factors cannot be estimated simultaneously. Methods to overcome these problems include regression procedures, such as forward selection (Westfall et al., 1998), stepwise and all-subsets regression (Abraham et al., 1999), and shrinkage methods, including the Smoothly Clipped Absolute Deviation procedure (Li and Lin, 2002) and the Dantzig selector (Phoa et al., 2009). We compare the performances of one representative from each of these classes of techniques, together with a new model-averaging procedure. Strategies are suggested for choosing values of the tuning constants for each analysis method. It is widely accepted that the effectiveness of supersaturated designs in detecting active factors requires there being only a small number of such factors, known as effect sparsity (Box and Meyer, 1986).

Previous simulation studies compared either a small number of analysis methods (Li and Lin, 2003; Phoa et al., 2009) or different designs (Allen and Bernshteyn, 2003), usually for a narrow range of settings. In our simulations, several settings are explored with different numbers and sizes of active effects, and a variety of design sizes. The results lead to guidance on when supersaturated designs are effective screening tools.

In Section 2 we describe the design criteria and model selection methods investigated in the simulation studies. Section 3 describes the studies and summarises the results. Finally, in Section 4, we discuss the most interesting findings and draw some conclusions about the effectiveness of the methods for different numbers and sizes of active effects.

2. Design criteria and model selection methods

We consider a linear main effects model for the response

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \tag{1}$$

where $\mathbf{Y}$ is the $n \times 1$ response vector, $\mathbf{X}$ is an $n \times (m+1)$ model matrix, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}$ is a vector of independent normally distributed random errors with mean 0 and variance σ^2. We assume that each of the m factors has two levels, ± 1. The first column of $\mathbf{X}$ is $\mathbf{1}_n = [1, \ldots, 1]^{\mathrm{T}}$, with column i corresponding to the levels of the $(i-1)$th factor $(i = 2, \ldots, m+1)$.

2.1. Design construction criteria

2.1.1. $E(s^2)$-optimality Booth and Cox (1962) proposed a criterion that selects a design by minimising the sum of the squared inner-products between columns i and j of $\mathbf{X}$ $(i, j = 2, \ldots, m+1;\ i \neq j)$. We extend this definition to include the inner-product of the first column with every other column of $\mathbf{X}$ to give

$$E(s^2) = \frac{2}{m(m+1)} \sum_{i<j} s_{ij}^2\,, \tag{2}$$

where s_{ij} is the ijth element of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ $(i, j = 1, \ldots, m+1)$. The two definitions are equivalent for balanced designs, that is, where each factor is set to +1 and -1 equally often. The balanced $E(s^2)$-optimal designs used in this paper were found using the algorithm of Ryan and Bulutoglu (2007). These designs achieve the lower bound on $E(s^2)$ for balanced designs given by these authors and, where more than one design satisfies the bound, a secondary criterion of minimising $\max_{i<j} s_{ij}^2$ is employed.

2.1.2. Bayesian D-optimality Under a Bayesian paradigm with conjugate prior distributions for $\boldsymbol{\beta}$ and σ^2 (O'Hagan and Forster, 2004, ch. 11), the posterior variance-covariance matrix for $\boldsymbol{\beta}$ is proportional to $(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{K}/\tau^2)^{-1}$. Here, $\tau^2\mathbf{K}^{-1}$ is the prior variance-covariance matrix for $\boldsymbol{\beta}$. Jones et al. (2008) suggested finding a supersaturated design that maximises

$$\phi_D = |\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{K}/\tau^2|^{1/(m+1)}\,.$$

They regarded the intercept β_0 as a *primary* term with large prior variance, and $\beta_1, \ldots, \beta_m$ as *potential* terms with small prior variances, see DuMouchel and Jones (1994), and set

$$\mathbf{K} = \begin{pmatrix} 0 & \mathbf{0}_{1\times m} \\ \mathbf{0}_{m\times 1} & \mathbf{I}_{m\times m} \end{pmatrix} . \tag{3}$$

The prior information can be viewed as equivalent to having sufficient additional runs to allow estimation of all factor effects. This method can generate supersaturated designs for any design size and any number of factors.

Bayesian D-optimal designs may be generated using a coordinate-exchange algorithm (Meyer and Nachtsheim, 1995). The value of τ^2 reflects the quantity of prior information; $\tau^2 = 1$ was used to obtain the designs presented. An assessment (not shown) of designs found for $\tau^2 = 0.2$ and $\tau^2 = 5$ indicated insensitivity of design performance to τ^2, see also Jones et al. (2008).

2.2. Model selection methods

Three methods are examined: regression (forward selection), shrinkage (Gauss-Dantzig selector), and model-averaging.

2.2.1. Forward selection This procedure starts with the null model and adds the most significant factor main effect at each step according to an F-test (Miller, 2002, pp. 39-42). The process continues until the model is saturated or no further factors are significant. The evidence required for the entry of a variable is controlled by the "F-to-enter" level, denoted by $\alpha \in (0, 1)$.

2.2.2. Gauss-Dantzig selector Shrinkage methods form a class of continuous variable selection techniques where each coefficient β_i is shrunk towards zero at a different rate. We investigate the Dantzig selector, proposed by Candes and Tao (2007), in which the estimator $\hat{\boldsymbol{\beta}}$ is the solution to

$$\min_{\hat{\boldsymbol{\beta}}\in\mathbb{R}^k} ||\hat{\boldsymbol{\beta}}||_1 \quad \text{subject to} \quad ||\mathbf{X}^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})||_\infty \leq \delta \, . \tag{4}$$

Here $||\boldsymbol{\beta}||_1 = |\beta_0| + \ldots + |\beta_m|$ is the l_1 norm, $||\boldsymbol{a}||_\infty = \max(|a_0|, \ldots, |a_m|)$ is the l_∞ norm, and δ is a tuning constant. The Dantzig selector essentially finds the most parsimonious estimator amongst all those that agree with the data. Optimisation (4) may be reformulated as a linear program and solved, for

example, using the package `lpSolve` (Berkelaar, 2007) in `R` (R Development Core Team, 2009).

Candes and Tao (2007) also developed a two-stage estimation approach, the Gauss-Dantzig selector, which reduces underestimation bias and was used for the analysis of supersaturated designs by Phoa et al. (2009). First the Dantzig selector is used to identify the active factors, and those factors whose coefficient estimates are greater than γ are retained. Second, least squares estimates are found by regressing the response on the set of retained factors.

2.2.3. Model-averaging Here inference is based on a subset of models rather than on a single model. For example, model-averaged coefficients are obtained by calculating estimates for a set of models and then computing a weighted average where the weights represent the plausibility of each model (Burnham and Anderson, 2002, ch. 4). This approach provides more stable inference under repeated sampling from the same process.

For a supersaturated design, it is often not computationally feasible to include all possible models in the procedure. Further, many models will be scientifically implausible and therefore should be excluded (Madigan and Raftery, 1994). Effect sparsity suggests restriction to a set of models each of which contains only a few factors. We propose a new iterative approach, motivated by the *many-models* method of Holcomb et al. (2007):

1. Fit all models composed of two factors and the intercept and calculate for each the value of the Bayesian Information Criterion (BIC)

$$BIC = n \log\Big(\frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n}\Big) + p \log(n)\,, \tag{5}$$

   where p is the number of model terms.

2. For model i, calculate a weight

$$w_i = \frac{\exp(0.5 \times \Delta BIC_i)}{\sum_{k=1}^{K} \exp(0.5 \times \Delta BIC_k)}\,, \qquad i = 1, \ldots, K\,,$$

   where $\Delta BIC_i = BIC_i - \min_{1,\ldots,K}(BIC_k)$ and $K = m(m-1)/2$.

3. For each factor, sum the weights of those models containing the factor. Retain the $m_1 < m$ factors with the highest summed weights. Parameter m_1 should be set fairly high to avoid discarding active factors.

4. Fit all possible models composed of three of the m_1 factors and the intercept. Calculate weights as in step 2. Retain the best $m_2 < m_1$ factors, as in step 3, to eliminate models of low weight and obtain more reliable inference.

5. Fit all M models composed of $m_3 < m_2$ factors and the intercept, where $M = m_2!/m_3!(m_2 - m_3)!$. Calculate new weights as in step 2.

6. Let $\beta^\star_{1r}, \ldots, \beta^\star_{m_2 r}$ be the coefficients of the m_2 factors in the rth model $(r = 1, \ldots, M)$, where we set $\beta^\star_{lr} = 0$ if the lth factor is not included in model r. Calculate model-averaged coefficient estimates

$$\bar{\beta}^\star_l = \sum_{r=1}^{M} w_r \hat{\beta}^\star_{lr} \,,$$

   where $\hat{\beta}^\star_{lr}$ is the least squares estimate of $\beta^\star_{lr}$ if factor l is in model r, and 0 otherwise.

7. Use an approximate t-test, on $n - m_3 - 1$ degrees of freedom, to decide if each of the m_2 factors is active. The test statistic is given by $\bar{\beta}^\star_l / \{\widehat{\mathrm{Var}}(\bar{\beta}^\star_l)\}^{1/2}$, where estimation of the model-averaged variance is discussed by Burnham and Anderson (2002, pp. 158-164).

The effectiveness of the each of the three methods described above depends on the values chosen for the tuning constants, discussed in Section 3.3.

3. Simulation Study and Results

We identified a variety of features of a typical screening experiment and combined these to provide settings of varying difficulty on which to test the design and model selection methods.

3.1. Features varied in the simulation

- *Ratio of factors to runs in the experiment.* Three choices of increasing difficulty were used and coded m_n: 22 factors in 18 runs (22_18), 24 in 14 (24_14) and 26 in 12 (26_12).

- *Design construction criteria.* To investigate the use of $E(s^2)$-optimal and Bayesian D-optimal designs, one design was found for each m_n under each criterion. These designs were then used for all simulations with m factors and n runs. For each design, the values of the objective functions $E(s^2)$ and ϕ_D are given in Table 1, together with the maximum (ρ_{max}) and minimum (ρ_{min}) correlations between factor columns.

  For each m_n, the designs have similar values of $E(s^2)$ and ϕ_D but different structures. The $E(s^2)$-optimal designs are balanced, whereas the Bayesian D-optimal designs have 9, 7, and 5 unbalanced columns for the 22_18, 24_14 and 26_12 experiments respectively, with column sums of ± 2. Also, the Bayesian D-optimal designs have a wider range of column correlations than the $E(s^2)$-optimal designs. In particular, ρ_{max} for an $E(s^2)$-optimal design is always less than or equal to that of the corresponding Bayesian D-optimal design.

- *Number and sizes of active factors.* The magnitude of the coefficient for each of the c active factors was drawn at random from a $N(\mu, 0.2)$ for the following scenarios:

  1. Effect sparsity: $c = 3$, $\mu = 5$.

Table 1: Values of objective functions and maximum and minimum column correlations for $E(s^2)$-optimal and Bayesian D-optimal designs used in the simulation study

Experiment	22_18		24_14		26_12	
Construction Criterion	$E(s^2)$	D	$E(s^2)$	D	$E(s^2)$	D
$E(s^2)$	5.3	5.4	7.2	7.1	7.5	7.3
ϕ_D	11.7	11.7	6.1	6.1	4.3	4.3
ρ_{max}	0.33	0.33	0.43	0.58	0.33	0.67
ρ_{min}	0.11	0	0.14	0	0	0

2. Intermediate complexity: $c = 4$ or $c = 5$ (chosen with equal probability) and $\mu = 4$.
3. Larger number of small effects: $c = 6$ and $\mu = 3$.
4. Larger number of effects of mixed size: $c = 9$ and one factor with each of $\mu = 10$, $\mu = 8$, $\mu = 5$, $\mu = 3$, and five factors with $\mu = 2$.

- *Model selection methods.* The four methods of Section 2.2 were applied and tuning constants chosen as described in Section 3.3.

3.2. Experiment simulation

For each of 10,000 iterations:

1. From columns $2, \ldots, m + 1$ of $\mathbf{X}$, c columns were assigned to active factors at random.
2. To obtain the coefficients for the active factors, a sample of size c was drawn from a $N(\mu, 0.2)$, and $\pm$ signs randomly allocated to each number.
3. Coefficients for the inactive factors were obtained as a random draw from a $N(0, 0.2)$.
4. Data were generated from model (1), with errors randomly drawn from a $N(0, 1)$, and analysed by each of the three model selection methods.

The random assignment of active factors to columns is important to remove selection bias. The choice of distributions at steps 2 and 3 ensures separation between the realised coefficients of the active and inactive factors.

3.3. Choice of tuning constants

For each method, a comparison of different values for the tuning constants was carried out prior to the main simulation studies. The aim was to find values of the tuning parameters that did not rely on detailed information from each simulation setting. This was achieved either by choosing values to give robust performance across the different settings, or by applying automated adaptive procedures.

Our strategy for the selection of δ and γ for the Gauss-Dantzig selector was to control type II errors via δ, by choosing a larger than necessary model with the Dantzig selector, and then control type I errors by choosing γ sufficiently large to screen for spurious effects. To choose δ, we used the standard BIC statistic (5) which gave similar results to the use of AIC. Phoa et al. (2009) proposed a modified AIC criterion which, in our study, consistently selected too few active effects when $c = 6$. The value of γ needs to be sufficiently small so that few active factors are declared inactive, but large enough for effects retained by the Dantzig selector to be distinguishable from the random error. This was achieved by the choice $\gamma = 1.5$.

Model-averaging is the most computationally demanding of the methods due to the large number of regression models fitted. In the choice of m_1, m_2 and m_3, a balance must be struck between discarding potentially active factors too early in the procedure, and including unlikely (for example, too large) models in the final step. Preliminary studies showed that $m_1 = 18$, $m_2 = 13$ and $m_3 = 8$ was an effective compromise. In step 5 of the procedure, some models may not be estimable. We found that removing a single factor overcame this problem. We therefore chose to remove the factor with smallest weight that produced a non-singular information matrix. Reassuringly, the power of the procedure to detect active effects (see Section 3.4) is relatively robust to the values of m_1 and m_2. Attempting to fit too large models in step 5, i.e. setting m_3 too high, can result in loss of power and also higher type I errors. We suggest that m_3 is chosen broadly in line with effect sparsity, and a little larger than the anticipated number of active factors.

In forward selection and model-averaging, $\alpha = 0.05$ was used based on investigations (not presented) that showed $\alpha > 0.05$ gave a substantial increase in type I errors without a corresponding increase in power. Decreasing α resulted in unacceptably low power for even the easiest simulation settings.

For each method studied, the results of the analysis can depend critically on the choice of tuning constants. The Gauss-Dantzig selector has the advantages of having a robust automated procedure for the choice of δ, and a straightfoward interpretation of γ as the minimum size of an active effect. This quantity may often be elicited from subject experts (see, for example, Δ, in Lewis and Dean, 2001).

Table 2: Simulation study results for 22_18 designs. FS=Forward Selection, GDS=Gauss-Dantzig Selector, MA=Model-Averaging; π_1=power, π_2=type I error rate, π_3=coverage, π_4=number of factors declared active

Design	$E(s^2)$-optimal			Bayesian D-optimal		
Analysis	FS	GDS	MA	FS	GDS	MA
Scenario 1: $c = 3$, $\mu = 5$						
π_1	1.00	1.00	1.00	1.00	1.00	1.00
π_2	0.11	0.01	0.02	0.11	0.03	0.02
π_3	1.00	1.00	1.00	1.00	1.00	1.00
π_4	5.07	3.17	3.41	5.07	3.61	3.42
Scenario 2: $c = 4$ or $c = 5$, $\mu = 4$						
π_1	0.89	1.00	0.99	0.90	1.00	0.99
π_2	0.09	0.01	0.02	0.10	0.04	0.02
π_3	0.85	0.99	0.98	0.86	0.99	0.98
π_4	5.57	4.72	4.74	5.69	5.11	4.74
Scenario 3: $c = 6$, $\mu = 3$						
π_1	0.57	0.93	0.90	0.58	0.95	0.89
π_2	0.06	0.02	0.02	0.06	0.04	0.02
π_3	0.37	0.77	0.74	0.38	0.82	0.73
π_4	4.35	5.97	5.63	4.43	6.26	5.62
Scenario 4: $c = 9$, $\mu = 10, 8, 5, 3, 2$						
π_1	0.56	0.73	0.55	0.56	0.75	0.56
$\pi_1(10, 8)$	1.00	1.00	1.00	1.00	1.00	1.00
$\pi_1(5, 3)$	0.78	0.92	0.80	0.78	0.93	0.80
$\pi_1(2)$	0.30	0.55	0.28	0.30	0.58	0.28
π_2	0.04	0.06	0.02	0.05	0.07	0.02
π_3	0.04	0.08	0.00	0.05	0.10	0.00
π_4	5.63	7.37	5.22	5.64	7.73	5.24

3.4. Simulation results

A factorial set of 54 simulations was run on the four features of Section 3.1. Four different criteria were used to assess performance of the designs and analysis methods:

Table 3: Simulation study results for 24_14 designs. FS=Forward Selection, GDS=Gauss-Dantzig Selector, MA=Model-Averaging; π_1=power, π_2=type I error rate, π_3=coverage, π_4=number of factors declared active

Design	$E(s^2)$-optimal			Bayesian D-optimal		
Analysis	FS	GDS	MA	FS	GDS	MA
Scenario 1: $c = 3$, $\mu = 5$						
π_1	0.86	0.98	0.91	0.86	0.99	0.90
π_2	0.11	0.03	0.05	0.11	0.04	0.05
π_3	0.82	0.97	0.84	0.82	0.98	0.81
π_4	4.77	3.54	3.83	4.81	3.75	3.83
Scenario 2: $c = 4$ or $c = 5$, $\mu = 4$						
π_1	0.53	0.85	0.73	0.53	0.89	0.72
π_2	0.09	0.06	0.06	0.09	0.07	0.06
π_3	0.31	0.69	0.50	0.31	0.76	0.48
π_4	4.03	5.04	4.47	4.02	5.25	4.45
Scenario 3: $c = 6$, $\mu = 3$						
π_1	0.31	0.61	0.46	0.30	0.65	0.46
π_2	0.08	0.09	0.09	0.08	0.09	0.09
π_3	0.02	0.20	0.11	0.02	0.26	0.10
π_4	3.22	5.23	4.29	3.16	5.57	4.31
Scenario 4: $c = 9$, $\mu = 10, 8, 5, 3, 2$						
π_1	0.40	0.53	0.40	0.40	0.56	0.39
$\pi_1(10, 8)$	0.90	0.98	0.90	0.91	0.99	0.88
$\pi_1(5, 3)$	0.47	0.65	0.45	0.49	0.69	0.45
$\pi_1(2)$	0.16	0.31	0.18	0.16	0.33	0.18
π_2	0.08	0.14	0.09	0.08	0.14	0.09
π_3	0.00	0.00	0.00	0.00	0.00	0.00
π_4	4.76	6.89	4.86	4.74	7.14	4.86

π_1**:** Average proportion of active factors correctly identified (Power; larger-the-better); for Scenario 4, the power was calculated separately for effects with $\mu = 10, 8$ (dominant; $\pi_1(10, 8)$), $\mu = 5, 3$ (moderate; $\pi_1(5, 3)$) and $\mu = 2$ (small; $\pi_1(2)$).

π_2**:** Average proportion of factors declared active which are actually inactive

Table 4: Simulation study results for 26_12 designs. FS=Forward Selection, GDS=Gauss-Dantzig Selector, MA=Model-Averaging; π_1=power, π_2=type I error rate, π_3=coverage, π_4=number of factors declared active

Design	$E(s^2)$-optimal			Bayesian D-optimal		
Analysis	FS	GDS	MA	FS	GDS	MA
Scenario 1: $c = 3$, $\mu = 5$						
π_1	0.66	0.89	0.67	0.67	0.92	0.68
π_2	0.11	0.06	0.09	0.11	0.06	0.10
π_3	0.54	0.82	0.48	0.56	0.87	0.47
π_4	4.43	3.95	4.14	4.44	4.10	4.36
Scenario 2: $c = 4$ or $c = 5$, $\mu = 4$						
π_1	0.37	0.65	0.45	0.36	0.69	0.47
π_2	0.10	0.10	0.11	0.10	0.10	0.12
π_3	0.10	0.35	0.15	0.09	0.41	0.15
π_4	3.71	5.00	4.36	3.63	5.22	4.62
Scenario 3: $c = 6$, $\mu = 3$						
π_1	0.25	0.43	0.31	0.25	0.47	0.32
π_2	0.10	0.11	0.12	0.09	0.11	0.13
π_3	0.00	0.04	0.01	0.00	0.07	0.01
π_4	3.47	4.71	4.31	3.36	5.00	4.46
Scenario 4: $c = 9$, $\mu = 10, 8, 5, 3, 2$						
π_1	0.31	0.44	0.31	0.31	0.47	0.32
$\pi_1(10, 8)$	0.75	0.93	0.68	0.75	0.94	0.69
$\pi_1(5, 3)$	0.31	0.48	0.31	0.31	0.52	0.32
$\pi_1(2)$	0.14	0.23	0.16	0.14	0.25	0.17
π_2	0.10	0.15	0.12	0.10	0.15	0.13
π_3	0.00	0.00	0.00	0.00	0.00	0.00
π_4	4.48	6.56	4.83	4.45	6.78	5.04

(Type I error rate; smaller-the-better).

π_3: Average proportion of simulations in which the set of factors declared active included all those truly active (Coverage; larger-the-better).

π_4: Average number of declared active factors.

The results are summarised in Tables 2, 3 and 4 for experiments 22_18, 24_14 and 26_12 respectively. These show that the Gauss-Dantzig selector has values of π_1 and π_3 as high, or higher, than the other analysis methods in almost all the simulations and often has very low values for π_2. The Gauss-Dantzig selector was found to have the most consistent performance of the three methods as measured by the variances (not shown) of the proportions involved in π_1, π_2 and π_3.

For the 22_18 experiment (Table 2), the performance of the Gauss-Dantzig selector is almost matched by the model-averaging method for Scenarios 1–3. However, the good performance of model-averaging is not maintained for the more difficult 24_14 and 26_12 experiments. The addition of extra steps in the procedure, such as fitting all four-factor models, may improve performance for larger numbers of factors at the cost of more computation.

Forward selection has consistently the worst performance measured by π_1 and π_3, and also performs poorly under π_2 for $c = 3$ and $c = 4, 5$. Also, the type I error rate (π_2), is always higher than the value set for the entry of a variable, $\alpha = 0.05$, due to the multiple testing.

From our comparisons, the Gauss-Dantzig selector is the most promising method, particularly in the more challenging settings. For example, for the 26_12 experiment with $c = 3$, $\mu = 5$ (Table 4), the Gauss-Dantzig selector has a 28% increase in power (π_1) and a 67% increase in coverage (π_3), relative to the next best method (model-averaging). In Scenario 4 for the 26_12 experiment, the Gauss-Dantzig selector performs extremely well in identifying the dominant factors ($\mu = 8$, $\mu = 10$), having $\pi_1 > 0.93$; the next best method (forward selection) has $\pi_1 = 0.75$. All methods had difficulty identifying small active effects.

To compare the supersaturated designs, we now focus on the preferred analysis via the Gauss-Dantzig selector. The Bayesian D-optimal designs have consistently higher values for $\pi_1, \ldots, \pi_4$ than the $E(s^2)$-optimal designs, although the differences are often small. Thus the Bayesian D-optimal designs lead to identifying as active a slightly higher proportion of both active and inactive effects. All designs give their best results for Scenario 1 and the worst for Scenarios 3 and 4, as expected. Performance deteriorates as the ratio of factors to runs increases, although good results are obtained under Scenario 1 (effect sparsity) for all m_n.

When there are few active factors in the 26_12 experiment, both classes of designs performed well under effect sparsity (Scenario 1, Table 4). To investigate if performance is maintained when the number of factors is increased

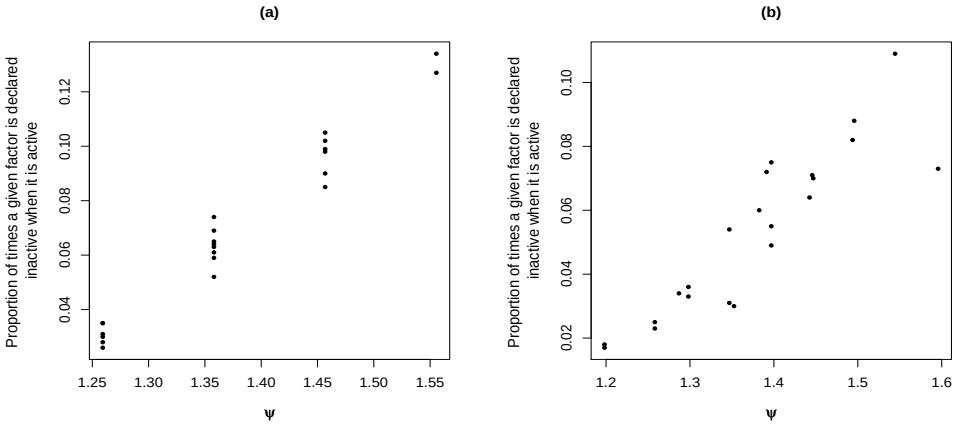

Fig. 1: Proportion of times a given factor was wrongly declared inactive plotted against ψ, for the 22_18 experiment and Scenario 3 analysed using the Gauss-Dantzig selector: (a) $E(s^2)$-optimal design; (b) Bayesian D-optimal design.

substantially, simulations of a 48_12 experiment were performed using each type of design. The results indicated poor performance with π_1 and π_3 less than 0.61 and 0.37 respectively.

In practice, the assignment of active factors to the columns of a design may influence the subsequent model selection. This was investigated by measuring the overall level of correlation of a given column j of $\mathbf{X}$ by

$$\psi_j = \sum_{i=2}^{m+1} \rho_{ij}^2 \,,$$

where ρ_{ij} is the correlation between columns i and j of $\mathbf{X}$ $(i, j = 2, \ldots, m+1)$. Fig. 1 shows the proportion of times that a given factor was wrongly declared inactive as a function of ψ for the 22_18 experiment and $c = 6$, $\mu = 3$, analysed using the Gauss-Dantzig selector. There are strong positive correlations for both the $E(s^2)$-optimal and Bayesian D-optimal designs, 0.98 and 0.90 respectively. Similar trends were observed for other simulated experiments and scenarios (not shown). This demonstrates the importance of using any prior information on the likely activity of factors when assigning them to columns of the design. For the Bayesian D-optimal design, any such information should ideally be incorporated in the design construction through

Table 5: Simulation results when there were no active factors. FS=Forward Selection, GDS=Gauss-Dantzig Selector, MA=Model-Averaging, SVD=SVDPRM; π_2=type I error rate, π_4=number of factors declared active

Design	$E(s^2)$-optimal			Bayesian D-optimal		
Analysis	FS	GDS	MA	FS	GDS	MA
22_18						
π_2	0.11	0.01	0.05	0.12	0.03	0.05
π_4	2.52	0.12	1.13	2.55	0.57	1.14
24_14						
π_2	0.12	0.01	0.08	0.12	0.02	0.08
π_4	2.88	0.23	1.85	2.82	0.43	1.88
26_12						
π_2	0.13	0.01	0.10	0.12	0.02	0.11
π_4	3.28	0.33	2.64	3.22	0.52	2.83

adjusting the elements of the prior dispersion matrix $\mathbf{K}$ in (3).

3.5. No active factors

Further simulations were used to check the performance of the design and analysis methods when there are no active factors, a situation where π_1 (power) and π_3 (coverage) no longer apply. From Table 5, the Gauss-Dantzig selector is clearly the best analysis method and rarely declares any factors active. The other methods have considerably higher type I errors, typically declaring at least two factors active. Table 5 also shows that the $E(s^2)$-optimal designs perform better than the Bayesian D-optimal designs for the Gauss-Dantzig selector, agreeing with the results for π_2 in Section 3.4.

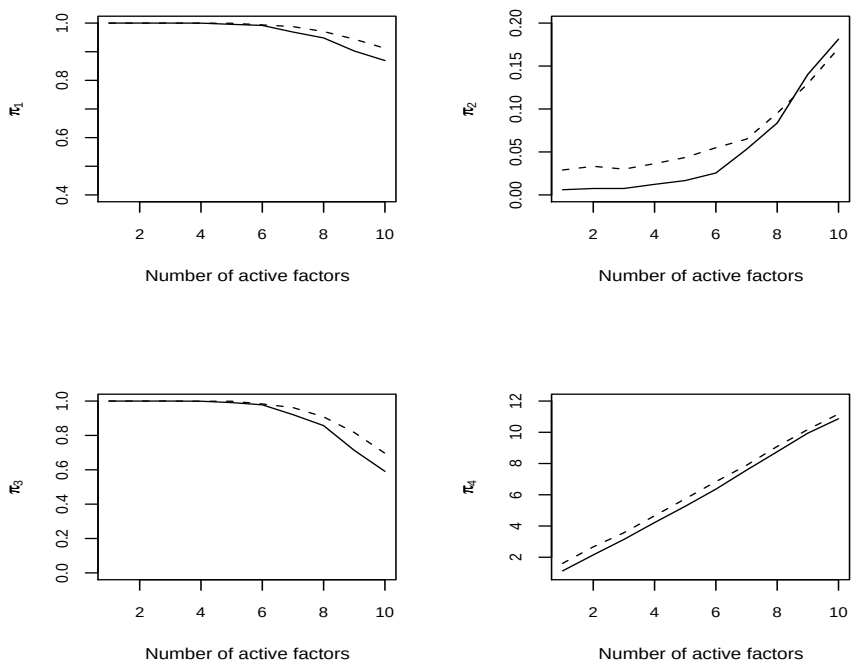

Fig. 2: Performance measures, $\pi_1, \ldots, \pi_4$, for the 22_18 experiment with $\mu = 5$ using the Gauss-Dantzig selector for $E(s^2)$ (solid line) and Bayesian D-optimal (dashed line) designs.

3.6. What is 'effect sparsity'?

A set of simulations was performed to assess how many active factors could be identified reliably using supersaturated designs. These simulations kept the mean, μ, of an active factor constant and varied the number of active factors, $c = 1, \ldots, 10$. Fig. 2 shows the four performance measures for the 22_18 experiment with $\mu = 5$ using the Gauss-Dantzig selector for analysis. Both the $E(s^2)$-optimal and the Bayesian D-optimal designs perform well for up to 8 active factors. The Bayesian D-optimal design has slightly higher π_1, π_2 and π_3 values and thus tends to select slightly larger models.

Fig. 3 shows the corresponding results for 24_14 experiment with $\mu = 3$. The performance, particularly under π_1 and π_3, declines more rapidly as the number of active factors increases. Again, slightly larger models are selected

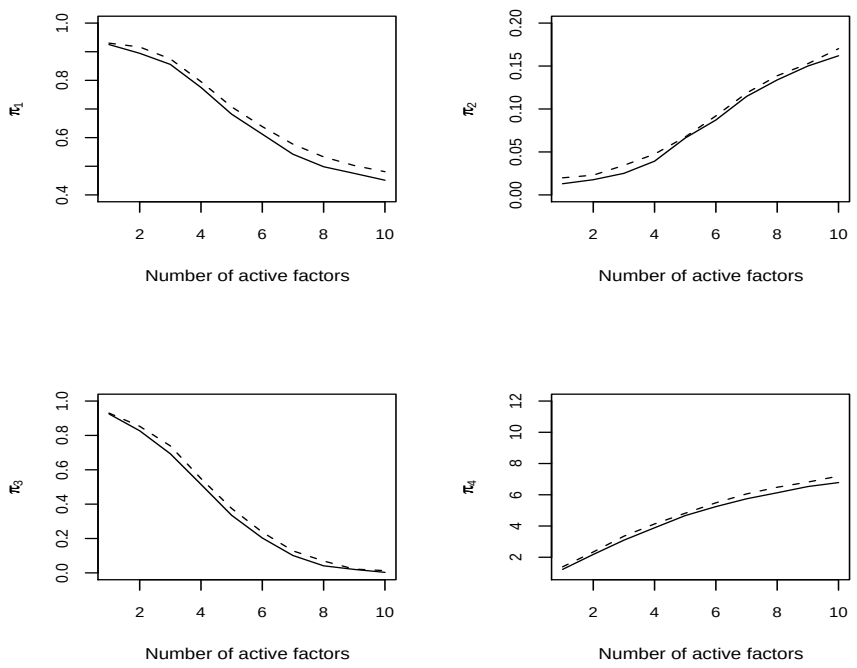

Fig. 3: Performance measures, $\pi_1, \ldots, \pi_4$, for the 24_14 experiment with $\mu = 3$ using the Gauss-Dantzig selector for $E(s^2)$ (solid line) and Bayesian D-optimal (dashed line) designs.

using the Bayesian D-optimal design, a difference which is not observed for other analysis methods. Further simulations (not shown) indicate considerable differences in performance between settings where $\mu = 3$ and $\mu = 5$.

4. Discussion

The results in this paper provide evidence that supersaturated designs may be a useful tool for screening experiments, particularly marginally supersaturated designs (where m is only slightly larger than n). They suggest the following guidelines for the use of supersaturated designs:

1. The Gauss-Dantzig selector is the preferred model selection procedure

out of the methods investigated. If the design is only marginally supersaturated, model-averaging is also effective.

2. The ratio of factors to runs should be less than 2.

3. The number of active factors should be less than or equal to a third of the number of runs.

The simulations include situations where these conditions do not hold but nevertheless a supersaturated design performs well, for example, Table 4 Scenario 1 with $m/n > 2$. However, evidence from our study suggests that 2 and 3 are conditions under which supersaturated designs are most likely to be successful.

Little difference was found in the performance of the $E(s^2)$-optimal and Bayesian D-optimal designs, with the latter having slightly higher power to detect active effects at the cost of a slightly higher type I error rate. The Bayesian D-optimal designs may be preferred in practice, despite being unbalanced and having some high column correlations, as follow-up experimentation may screen out spurious factors but cannot detect active factors already removed. Such designs are readily available in standard software such as SAS Proc Optex and JMP.

The simulations presented cover a broader range of conditions than previously considered, and investigate more aspects of design performance. Further studies of interest include incorporating interaction effects in the models, and Bayesian methods of analysis, see for example Beattie et al. (2002).

Acknowledgments

The authors wish to thank Bradley Jones (SAS) and Susan Lewis (University of Southampton) for helpful discussions. The first author was supported by a PhD studentship from EPSRC and the Lubrizol Corporation.

References

Abraham, B., Chipman, H., Vijayan, K., 1999. Some risks in the construction and analysis of supersaturated designs. Technometrics 41, 135–141.

Allen, T. T., Bernshteyn, M., 2003. Supersaturated designs that maximize the probability of identifying active factors. Technometrics 45, 90–97.

Beattie, S. D., Fong, D. K. H., Lin, D. K. J., 2002. A two-stage Bayesian model selection strategy for supersaturated designs. Technometrics 44, 55–63.

Berkelaar, M., 2007. lpSolve: Interface to Lp_solve v. 5.5 to solve linear/integer programs.

Booth, K. H. V., Cox, D. R., 1962. Some systematic supersaturated designs. Technometrics 4, 489–495.

Box, G. E. P., 1959. Discussion of the papers of Satterthwaite and Budne. Technometrics 1, 174–180.

Box, G. E. P., Meyer, R. D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28, 11–18.

Burnham, K. P., Anderson, D. R., 2002. Model Selection and Multimodel Inference, 2nd Edition. Springer.

Candes, E., Tao, T., 2007. The Dantzig selector: Statistical estimation when p is much larger than n. The Annals of Statistics 35, 2313–2351.

DuMouchel, W., Jones, B., 1994. A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. Technometrics 36, 37–47.

Gilmour, S. G., 2006. Supersaturated designs in factor screening. In: Dean, A. M., Lewis, S. M. (Eds.), Screening. Springer, New York, pp. 169–190.

Holcomb, D. R., Montgomery, D. C., Carlyle, W. M., 2007. The use of supersaturated experiments in turbine engine development. Quality Engineering 19, 17–27.

Jones, B., Lin, D. K. J., Nachtsheim, C. J., 2008. Bayesian D-optimal supersaturated designs. Journal of Statistical Planning and Inference 138, 86–92.

Lewis, S. M., Dean, A. M., 2001. Dectection of interactions in experiments on large numbers of factors (with discussion). Journal of the Royal Statistical Society B 63, 633–672.

Li, R., Lin, D. K. J., 2002. Data analysis in supersaturated designs. Statistics & Probability Letters 59, 135–144.

Li, R., Lin, D. K. J., 2003. Analysis methods for supersaturated design: Some comparisons. Journal of Data Science 1, 249–260.

Li, W. W., Wu, C. F. J., 1997. Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. Technometrics 39, 171–179.

Lin, D. K. J., 1993. A new class of supersaturated designs. Technometrics 35, 28–31.

Lin, D. K. J., 1995. Generating systematic supersaturated designs. Technometrics 37, 213–225.

Madigan, D., Raftery, A. E., 1994. Model selection and accounting for model uncetainty in graphical models using Occam's window. Journal of the American Statistical Association 89, 1535–1546.

Meyer, R. K., Nachtsheim, C. J., 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. Technometrics 37, 60–69.

Miller, A., 2002. Subset Selection in Regression. Chapman and Hall, Boca Raton.

Nguyen, N. K., 1996. An algorithmic approach to constructing supersaturated designs. Technometrics 38, 69–73.

Nguyen, N. K., Cheng, C. S., 2008. New $E(s^2)$-optimal supersaturated designs constructed from incomplete block designs. Technometrics 50, 26–31.

O'Hagan, A., Forster, J. J., 2004. Kendall's Advanced Theory of Statistics Vol. 2B: Bayesian Inference, 2nd Edition. Arnold, London.

Phoa, F. K. H., Pan, Y.-H., Xu, H., 2009. Analysis of supersaturated designs via the Dantzig selector. Journal of Statistical Planning and Inference in press, doi:10.1016/j.jspi.2008.10.023.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ryan, K. J., Bulutoglu, D. A., 2007. $E(s^2)$-optimal supersaturated designs with good minimax properties. Journal of Statistical Planning and Inference 137, 2250–2262.

Satterthwaite, F., 1959. Random balance experimentation. Technometrics 1, 111–137.

Westfall, P. H., Young, S. S., Lin, D. K. J., 1998. Forward selection error control in the analysis of supersaturated designs. Statistica Sinica 8, 101–117.

Wu, C. F. J., 1993. Construction of supersaturated designs through partially aliased interactions. Biometrika 80, 661–669.