RR-95-03

A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing

H. Chipman and M. Hamada
University of Waterloo
C.F.J. Wu
University of Michigan

March 1995

A BAYESIAN VARIABLE SELECTION APPROACH FOR ANALYZING DESIGNED EXPERIMENT WITH COMPLEX ALIASING

H. Chipman
Graduate School of Business
University of Chicago
Chicago, IL 60637, USA

Michael Hamada
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON N2L 3G1, Canada

and

C.F.J. Wu
Department of Statistics
and of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109-1027, USA

ABSTRACT

Often experiments using designs with complex aliasing patterns are performed, e.g., two-level non-geometric Plackett-Burman designs, multi-level and mixed-level fractional factorial designs, two-level fractional factorial designs with hard-to-control factors and supersaturated designs. Hamada and Wu (1992) proposed an iterative guided stepwise regression strategy for analyzing the data from such designs that allowed the entertainment of interactions. Their strategy provides a restricted search in a rather large model space, however. This paper provides an efficient methodology for searching the model space more thoroughly and is based on a Bayesian variable selection algorithm. It is shown how the use of hierarchical priors provides a flexible and powerful way to focus the search on a reasonable class of models in the model space. The proposed methodology is demonstrated with four examples, three of which come from actual industrial experiments.

Key words: Gibbs Sampler, Hard-to-control Factors, Interactions, Partial Aliasing, Plackett-Burman Designs, Supersaturated Designs.

1 Introduction

Non-geometric (i.e., whose run sizes are not a power of two) Plackett-Burman (1946) (PB) designs such as those with 12, 20 and 24 runs are popular for screening a large number of two-level factors because of their run size economy. Traditionally, the analysis of these PB designs has been confined to main effects only under the assumption that the interactions are negligible. This focus on main effects is due to the complex aliasing patterns of these designs. Consider the 12-run PB design with 11 factors in Table 1: for each factor, say X, its main effect is partially aliased with the 45 two-factor interactions not involving X. Because of such complex aliasing patterns, Daniel (1976, page 294) had reservations about using PB designs even for screening and referred to their complex aliasing patterns as "hazards."

Hamada and Wu (1992) went beyond the traditional approach by showing that interactions could be identified and estimated with reasonable precision from such designs with complex aliasing. They proposed an iterative analysis strategy based on the precepts of *effect sparsity* (i.e., experimental variation is attributed to only a few effects) and *effect heredity* (i.e., a significant two-factor interaction occurs with at least one of the corresponding main effects being significant) which exploited the designs' complex aliasing patterns. Recognizing the potential for entertaining interactions, the "hazards" of the non-geometric PB designs could now be viewed as "advantages." For example, in geometric PB designs (i.e., 2^{k-p} fractional factorials), a main effect is either orthogonal to or completely aliased with an interaction so that if main effect A is completely aliased with interaction BC, the geometric PB design would provide no information about their separate effects; in contrast, for a non-geometric PB design, the two effects are partially aliased and can be decoupled under certain assumptions.

Designs and data with complex aliasing patterns arise in a number of situations and are listed below:

- **two-level factors**

  In addition to the non-geometric 12, 20, 24 and 28 PB designs (Plackett and Burman 1946), Hall (1966) gave four non-geometric 16 run designs.

- **multi-level and mixed level fractional factorials**

3^{k-p} fractional factorials are examples of multi-level designs in which complex aliasing arises if each main effect is decomposed into the linear and quadratic contrasts and each 2-factor interaction into linear × linear, linear × quadratic, quadratic × linear, and quadratic × quadratic contrasts. $L_{18}(2 \times 3^7)$ and $L_{36}(2^{11} \times 3^{12})$ are examples of mixed level designs which accommodate both two-level and three-level factors. See Wang and Wu (1991) for a large number of classes of mixed-level designs. Wang and Wu (1992) also considered "nearly orthogonal" designs whose main effects are either orthogonal or nearly orthogonal and which also have complex aliasing patterns.

- **hard-to-control factors**

  There may be some difficulty in controlling the experimental factors exactly so that the experimental design is not carried out as planned. Consequently, even a 2^{k-p} fractional factorial design will no longer be orthogonal when improperly implemented and therefore have complex aliasing patterns. Also, a mistake may be made in setting the factors levels for a particular run which will have the same adverse effect.

- **supersaturated designs**

  Supersaturated designs allow the study of more factors than runs. See some recent work by Lin (1993) and Wu (1993) which presented designs that have complex aliasing. In fact those given in Wu (1993) use the partially aliased interaction columns of the PB designs to accommodate the additional factors.

The analysis strategy in Hamada and Wu (1992) was motivated by the potential infeasibility of performing all-subsets regression with main effects and all two-factor interactions; namely,

(i) more effects than runs (or observations)

(ii) computationally infeasible, say with 66 effects for a 12-run PB design

(iii) potential unreasonable models with two-factor interactions and no main effects.

The Hamada and Wu (1992) analysis strategy used an iterative stepwise regression approach which addressed (i) and (ii) and was guided by the principle of *effect heredity* which addressed (iii). Note

that their strategy did not explicitly impose effect heredity, so models with two-factor interactions without corresponding main effects may still be obtained.

While providing a feasible alternative to an all-subsets regression, the Hamada and Wu (1992, p. 132) strategy limits the search in a rather large model space. In fact, Hamada and Wu (1992, p. 136) modified their original strategy to identify an additional class of models missed by the original strategy. Moreover, with more effects than runs, there is the possibility of several models explaining the data equally well which the investigator would want to know. The stepwise strategy tends to identify a single model, however. Consequently, there is a need for a feasible all-subsets strategy which addresses (i)-(iii). This is the thrust of the present paper which considers a more thorough search of the model space in a stochastic fashion and is based on the recent methodology reported in Chipman (1994). The stochastic search is accomplished using Bayesian variable selection methods. A suitable class of hierarchical prior distributions focuses the search in the model space on a reasonable class of models as suggested by (iii) above (i.e., that obey effect heredity). The stochastic nature of the search means that all models have positive probability of being visited. The fact that this stochastic search is data guided means two things for this procedure. First, if there is more than one model suggested by the data, the stochastic search will not get stuck at one model, because the data will guide the stochastic walk to both models. Second, since the data will typically suggest that a small subset of the model space is most likely, reasonable estimates of the probability of these models are available based on many fewer posterior samples than the total number of models.

The paper is organized as follows. In Section 2, four experiments (three with real data) are given which illustrate the situations in which complex aliasing arises. In Section 3, a Bayesian variable selection algorithm which incorporates the hierarchical model requirements, i.e., *Bayesian hierarchical model selection*, is presented. The experiments given in the previous section are analyzed in Section 4 using the Bayesian hierarchical model selection methodology. The paper concludes with a discussion in Section 5.

2 Examples

In this section, examples of four experiments illustrating situations in which complex aliasing arises are given. These include a screening experiment using a Plackett-Burman 12-run design, a mixed-level design, an experiment with hard-to-control factors and a supersaturated design.

2.1 Screening Experiment

Table 1 presents a 12-run PB design and illustrates its use in a screening context which can accommodate up to 11 factors labeled A-K. The data were originally constructed in Hamada and Wu (1992) based on the true model $Y = A + 2AB + 2AC + \epsilon$ with $\epsilon \sim N(0, \sigma = 0.25)$; i.e., factors A, B and C are active with the remaining factors $D - K$ inactive. For an actual experiment, see Hamada and Wu (1992), which reanalyzed a 12-run PB design to improve the reliability of weld repaired casts, originally due to Hunter, Hodi and Eager (1982)

Table 1: Screening experiment with Plackett-Burman 12-run design and response data

design											
A	B	C	D	E	F	G	H	I	J	K	response
+	+	−	+	+	+	−	−	−	+	−	1.058
+	−	+	+	+	−	−	−	+	−	+	1.004
−	+	+	+	−	−	−	+	−	+	+	-5.200
+	+	+	−	−	−	+	−	+	+	−	5.320
+	+	−	−	−	+	−	+	+	−	+	1.022
+	−	−	−	+	−	+	+	−	+	+	-2.471
−	−	−	+	−	+	+	−	+	+	+	2.809
−	−	+	−	+	+	−	+	+	+	−	-1.272
−	+	−	+	+	−	+	+	+	−	−	-0.955
+	−	+	+	−	+	+	+	−	−	−	0.644
−	+	+	−	+	+	+	−	−	−	+	-5.025
−	−	−	−	−	−	−	−	−	−	−	3.060

Table 2: Blood glucose experiment with mixed-level design and response data

design								mean
A	G	B	C	D	E	F	H	reading
1	1	1	1	1	1	1	1	97.94
1	1	2	2	2	2	2	2	83.40
1	1	3	3	3	3	3	3	95.88
1	2	1	1	2	2	3	3	88.86
1	2	2	2	3	3	1	1	106.58
1	2	3	3	1	1	2	2	89.57
1	3	1	2	1	3	2	3	91.98
1	3	2	3	2	1	3	1	98.41
1	3	3	1	3	2	1	2	87.56
2	1	1	3	3	2	2	1	88.11
2	1	2	1	1	3	3	2	83.81
2	1	3	2	2	1	1	3	98.27
2	2	1	2	3	1	3	2	115.52
2	2	2	3	1	2	1	3	94.89
2	2	3	1	2	3	2	1	94.70
2	3	1	3	2	3	1	2	121.62
2	3	2	1	3	1	2	3	93.86
2	3	3	2	1	2	3	1	96.10

2.2 Blood Glucose Experiment Using Mixed-Level Design

Henkin (1986) used an 18-run mixed-level design to study the effect of one two-level factor and seven three-level factors on blood glucose readings made by a clinical laboratory testing device. Note that all the factors were quantitative. Here we consider only one aspect of the study which was to identify factors that affect the mean reading. The design and response data are given in Table 2.

2.3 Experiment with Hard-to-Control Factors

The design given in Table 3 was used in a real experiment on a wood pulp production process which studied 11 factors. Quality characteristics such as yield, burst index and opacity were observed. The process consisted of chemical and mechanical treatments; factors 1-7 involve the chemical treatment, while factors 8-11 involve the mechanical treatment. The planned experiment was a

Plackett-Burman 20-run design with a center point replicated twice (i.e. the total run size was 22) . Data from only 19 runs were available since difficulties were encountered in performing three of the runs from the Plackett-Burman design portion. Also, notice that several of the factors were hard to control, notably, factors 5, 9 and 11 (wood to liquid ratio, slurry concentrations at two stages); the planned levels were ± 1 in runs 1-17 and 0 in runs 18-19. The actual factor levels and the observed quality characteristic, burst index, are given in Table 3.

Table 3: Experiment with hard-to-control factors, design and response data

design											
x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	response
1	-1.00	1	1	-0.33	-1	1	1	0.74	1	-0.89	1.61
-1	-1.00	1	1	1.63	1	-1	1	-1.02	1	-0.76	1.97
-1	0.99	-1	-1	-1.04	-1	1	1	-0.55	1	1.85	1.48
1	1.00	-1	1	1.82	-1	-1	1	0.35	1	1.03	0.55
1	1.17	-1	-1	0.31	1	1	1	-0.67	1	-1.08	0.55
-1	-1.00	-1	-1	1.00	1	-1	1	0.75	-1	-0.87	1.59
1	1.00	-1	1	-0.57	1	-1	-1	-1.19	-1	2.26	1.64
-1	-1.00	-1	-1	-0.32	-1	-1	-1	-1.16	-1	-0.79	1.50
1	1.00	1	-1	1.69	-1	1	-1	-1.20	-1	-0.87	1.97
-1	-1.00	-1	1	1.32	-1	1	1	-1.17	-1	2.17	1.67
1	-0.98	1	-1	1.57	-1	-1	-1	-1.41	1	1.12	1.52
-1	1.00	1	1	1.61	-1	1	-1	-0.77	-1	-0.40	4.37
-1	1.00	1	-1	-1.06	1	1	1	0.45	-1	2.32	2.38
1	1.00	1	1	-0.76	1	-1	1	-0.62	-1	-0.83	2.04
-1	-1.00	1	1	-0.33	1	1	-1	-1.69	1	-1.38	2.24
1	-1.00	-1	1	1.36	1	1	-1	3.35	-1	0.66	1.76
-1	1.00	-1	1	-0.23	-1	-1	-1	1.45	1	-0.65	1.73
0	0.00	0	0	-0.10	0	0	0	-0.09	0	0.39	1.74
0	0.00	0	0	0.05	0	0	0	0.58	0	0.16	1.76

2.4 Experiment with Supersaturated Design

Lin (1993) showed that a half-fraction of a Plackett-Burman design could be used as a supersaturated design. He illustrated this with a 28-run Plackett-Burman design with 24 factors from an experiment used to develop an epoxide adhesive system as reported by Williams (1968). The half

fraction (based on an unused orthogonal column, yielding runs 1, 3, 4, 6, 8-10, 13, 17, 22-25, 28) of the original design along with the corresponding strip adhesion response data are displayed in Table 4. This illustrates the use of a 14-run design to study 23 factors; note that in the half fraction factors 13 and 16 were assigned to the same column so that only factor 13 is reported here.

Table 4: Experiment with supersaturated design and response data

design																								
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	17	18	19	20	21	22	23	24	response
1	1	1	-1	-1	-1	1	1	1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	-1	1	133
1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	-1	-1	62
1	1	-1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	1	-1	-1	-1	-1	1	1	-1	45
1	1	-1	1	-1	1	-1	-1	-1	1	1	-1	1	-1	1	-1	1	1	1	-1	-1	-1	-1	52
-1	-1	1	1	1	1	-1	1	1	-1	-1	-1	1	-1	1	1	-1	-1	1	-1	1	1	1	56
-1	-1	1	1	1	1	1	-1	1	1	1	-1	-1	1	1	1	1	1	1	1	1	-1	-1	47
-1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	1	-1	1	1	1	1	1	-1	-1	1	88
-1	1	1	-1	-1	1	-1	1	-1	1	-1	-1	-1	-1	-1	-1	-1	1	-1	1	1	1	-1	193
-1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	32
1	1	1	1	-1	1	1	1	-1	-1	-1	1	-1	1	1	1	-1	1	-1	1	-1	-1	1	53
-1	1	-1	1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	276
1	-1	-1	-1	1	1	1	-1	1	1	1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	145
1	1	1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	-1	130
-1	-1	1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	-1	-1	127

3 Stochastic Variable Selection

This section reviews one algorithm for variable selection, based on the Gibbs sampler (see Smith and Roberts (1993) and references therein for an overview). The criterion of interest is taken to be the posterior probability of a model conditional on the data which can be obtained using the stochastic search variable selection (SSVS) algorithm of George and McCulloch (1993). The approach can be outlined as follows for the simplest case of linear regression with normal errors,

$$Y = X'\beta + \sigma\epsilon, \qquad \epsilon \sim N(0,1). \tag{1}$$

The central concept is to introduce an unobserved vector $\boldsymbol{\delta}$ of zeros and ones of length p, the same length as $\boldsymbol{\beta}$. The components of this vector represent the importance of the corresponding predictor variables. That is, if $\delta_i = 0$, then the magnitude of β_i is small, and the corresponding predictor is "inactive". If $\delta_i = 1$, then the magnitude of β_i is large, and the predictor is "active". Mathematically, this is accomplished by defining a normal mixture prior for the coefficients $\boldsymbol{\beta}$:

$$f(\beta_i|\delta_i) = \begin{cases} N(0, \tau_i^2) & \text{if } \delta_i = 0 \\ N(0, (c_i\tau_i)^2) & \text{if } \delta_i = 1 \end{cases} \tag{2}$$

When $\delta_i = 0$, β_i is tightly centered around 0, and will not have a large effect. The much larger variance ($c_i >> 1$) when $\delta_i = 1$ allows the possibility of a variable having a large influence. Thus, the parameters τ_i and c_i must be chosen to represent respectively a "small" effect, and how many times larger a "large" effect should be.

Since $\boldsymbol{\delta}$ is unknown, a prior distribution is placed on it. The commonly used independence prior for $\boldsymbol{\delta}$ implies that the importance of any variable is independent of the importance of any other variable. This will not be the case here, since the importance of interactions can be assumed to depend on the importance of their corresponding main effects. Hierarchical priors for interactions and polynomial terms, developed by Chipman (1994) are used to formally express these relations in a flexible fashion. These priors are described in Section 3.1.

A prior must also be specified for σ; following George and McCulloch (1993), we take

$$\sigma^2 \sim \text{IG}(\nu/2, \nu\lambda/2),$$

where IG denotes an inverted gamma distribution. This is equivalent to $\nu\lambda/\sigma^2 \sim \chi^2_\nu$.

This specific parameterization is chosen so that a Gibbs sampling approach may be used to obtain the posterior for $\boldsymbol{\delta}$. The Gibbs sampler utilizes conditional distributions to produce a sequence of samples from the posterior distribution. Such a technique is useful when the posteriors are not available in closed form, which is the case here. Discussion of this technique in general is given in Smith and Roberts (1993); George and McCulloch (1993) and Chipman (1994) discussed its application to variable selection. Here, as in George and McCulloch (1993), the algorithm consists of a multivariate normal draw for $\boldsymbol{\beta}|\boldsymbol{\delta}, \sigma$, an inverse gamma draw for $\sigma|\boldsymbol{\beta}, \boldsymbol{\delta}$ and p Bernoulli draws for $\delta_i|\boldsymbol{\beta}, \sigma, \{\delta_j\}_{j\neq i}$.

3.1 Hierarchical Priors for Variable Selection

The idea behind the dependence relation is that the importance of a higher order term depends on those lower order terms from which it was formed. Consider a simple example in which there are three main effects A, B, C and three two-factor interactions AB, AC, and BC. The importance of, say, AB will depend on whether the main effects A and B are included in the model. If neither are, then the interaction seems less plausible, as well as being more difficult to explain. This belief can be expressed in the prior for $\boldsymbol{\delta} = (\delta_A, \ldots, \delta_{BC})$ as follows:

$$\Pr(\boldsymbol{\delta}) = \Pr(\delta_A)\Pr(\delta_B)\Pr(\delta_C)\Pr(\delta_{AB}|\delta_A,\delta_B)\Pr(\delta_{AC}|\delta_A,\delta_C)\Pr(\delta_{BC}|\delta_B,\delta_C). \tag{3}$$

In (3), two principles are used to obtain the simplified form. First, the *conditional independence principle* assumes that conditional on first order terms, the second order terms $(\delta_{AB}, \delta_{AC}, \delta_{BC})$ are independent. Independence is also assumed for main effects. The *inheritance principle* assumes that the importance of a term depends only on those terms from which it was formed, implying $\Pr(\delta_{AB}|\delta_A,\delta_B,\delta_C) = \Pr(\delta_{AB}|\delta_A,\delta_B)$.

The exact nature of this dependence on "parent" terms is defined by the components of the joint probability in (3). For example, the probability that the term AB is active $\Pr(\delta_{AB} = 1|\delta_A,\delta_B)$ takes on four different values:

$$P(\delta_{AB} = 1|\delta_A,\delta_B) = \begin{cases} p_{00} & \text{if } (\delta_A,\delta_B) = (0,0) \\ p_{01} & \text{if } (\delta_A,\delta_B) = (0,1) \\ p_{10} & \text{if } (\delta_A,\delta_B) = (1,0) \\ p_{11} & \text{if } (\delta_A,\delta_B) = (1,1) \end{cases}. \tag{4}$$

Here, we will choose p_{00} small (e.g., 0.01), p_{01} and p_{10} larger (e.g., 0.10) and p_{11} largest (e.g., 0.25). This represents the belief that a two-factor interaction without parents is quite unlikely, one with a single parent is more likely, and one with both parents is most likely. The term *relaxed weak heredity* will refer to this prior, and setting $p_{00} = 0$ yields *strict weak heredity*

The prior developed for two-factor interactions may be generalized to polynomials and interactions involving polynomials. Since the concepts are similar, we simply summarize the results here and refer the reader to Chipman (1994) for more details. Consider a simple example, with fourth order term A^2B^2, third order terms AB^2, A^2B, second order terms A^2, AB, B^2 and first order terms A and B. We consider the parents of a term to be those terms of the next smallest order which can

form the original term when multiplied by a main effect. We will assume that the importance of a term depends only on these parents, an assumption called the *immediate inheritance principle.* Here, A^2B^2 has parents A^2B (since multiplication by B produces A^2B^2) and AB^2. Some terms (such as A^2) will have only one parent (e.g., A). The principles of conditional independence (which now says that terms of a given order are independent given all lower order terms) and inheritance are again applied, with the immediate inheritance principle. Other forms of hierarchical dependence are possible; this specific definition was chosen for the examples in this paper.

Another interesting class of predictors is qualitative predictors such as treatment, supplier, or location. Although not present in the examples of this paper, such variables arise in screening experiments, often in the form of three-level factors. Dummy variables are commonly used in such a situation, but typically one wants either all or none of the variables to be included in the model. The prior introduced by Chipman (1994) for this situation maps a single element $\delta_i \in \boldsymbol{\delta}$ to all of the dummy variables associated with a single factor, forcing them to either all be in or out of the model. Extensions to interactions involving qualitative factors are straightforward.

3.2 Choice of Prior Parameters

Before analyzing the data, some prior parameters must be chosen. The normal mixture prior on $\boldsymbol{\beta}$ has parameters τ and c, and the inverse gamma prior for σ has parameters ν and λ. Since this methodology is used as a tool rather than strictly for Bayesian reasons, we view these parameters as tuning constants as well as representations of prior information.

As in Box and Meyer (1986), we use $c = 10$, which indicates that an important effect is an order of magnitude larger than an unimportant one. For the choice of τ, we take, as in George and McCulloch (1993),

$$\tau_j = \Delta Y / 3\Delta X_j, \tag{5}$$

where ΔY represents a "small" change in Y, and ΔX_j represents a "large" change in X_j. This implies that if $\delta_j = 0$, even a large change in X_j is unlikely to produce any more than a small change in Y. A value of ΔY still must be chosen. When expert opinions are not available,

$$\Delta Y = \sqrt{\mathrm{Var}(Y)}/5,$$

where $\mathrm{Var}(Y)$ is the sample variance of the response without any regression, is found to work well in practice. This choice corresponds to the belief that after a model is fit to the data, σ will be roughly 20% of the unadjusted standard deviation.

An improper prior (i.e., $\nu = 0$) for σ is not appropriate, since this allows unreasonably large σ values, which can lead to considerable mass on the null model, i.e., the model with only an intercept. For σ^2, we will work with priors on the root scale. Choosing a prior for σ which makes large values implausible corrects this problem. The assumption that $\sigma \approx \sqrt{\mathrm{Var}(Y)}/5$ suggests that a prior on σ with a mean equal to $\sqrt{\mathrm{Var}(Y)}/5$ be used. Among these priors, the desirable spread may be attained by selecting a prior with an upper quantile (say 99%) that is near $\mathrm{Var}(Y)$. This approach often yields a value of ν near 2, which corresponds to a reasonably uninformative prior. The value of λ changes from experiment to experiment, since it depends on the scale of the response measurement.

4 Analysis

The Bayesian hierarchical model selection methodology presented in the previous section will be illustrated using the four examples given in Section 2.

4.1 Screening Design Experiment

The data were originally constructed in Hamada and Wu (1992) to illustrate that their stepwise strategy for variable selection could have difficulties identifying interactions if the corresponding main effects were smaller. The stepwise nature of their procedure caused it to miss all three active terms, which suggests the proposed approach might be more effective.

Since the true model is known, we used several different priors to assess the influence of different parameters. Prior means for σ of (0.1,0.63,1.0) were used with a (ν, λ) pair chosen to give good coverage. The first and third priors are intended to represents "extreme" situations, while the middle value was chosen using the $\sqrt{\mathrm{Var}(Y)}$ rule. It was found that conclusions were quite similar, so only those for a prior mean of 0.63 of σ are reported.

Table 5: Screening experiment posterior model probabilities

model	$\tau^*/2$	τ^*	$2\tau^*$
A, AB, AC	0.103	0.325	0.094
A,C,AB,AC	0.016	0.039	0.009
A,B,AB,AC	0.018	0.022	0.008
I,DF	—	—	0.013
K,GJ	—	—	0.009

We also need to specify the probabilities that factors are active. A relaxed weak heredity prior will be used (see Section 3.1), with

$$\Pr(\delta_A = 1) = 0.25$$

$$\Pr(\delta_{AB} = 1) = \begin{cases} 0.01 & \text{if } \delta_A = \delta_B = 0 \\ 0.10 & \text{if } \delta_A \neq \delta_B \\ 0.25 & \text{if } \delta_A = \delta_B = 1 \end{cases} .$$

This prior allows interactions to be active if only one parent term is active, and even if both parents are inactive, there is a small probability that the interaction will be active.

The parameter τ was also studied. The estimate based on the rule of thumb (5) is $\tau^* = 0.103$. To examine the robustness of conclusions to τ, we performed analyses with $\tau^*/2, \tau^*$, and $2\tau^*$.

The Gibbs algorithm was run 10,000 times, and every second sample used. Examination of autocorrelations and output from several independent runs confirmed that convergence and mixing occurred quickly, and that the suggested sample size is adequate. Table 5 and Figure 1 give joint and marginal posteriors for three different values of τ. Although the correct model has the most mass in all three cases, there is considerable dependence on τ. When a small value is used ($\tau^*/2$), too many effects are considered "large", leading to less model certainty. As τ increases to τ^*, there is less model uncertainty, and the correct model receives the most mass. As τ increases further to $2\tau^*$, there is again less model certainty. The behavior of the algorithm for large τ values is better understood by looking at the marginal distributions. From them, it is clear that no term appears to be important, as one would expect when τ is too large. The correlations between the many candidate terms result in many models receiving some posterior mass, many of them nonsensical.

In this case, the method works quite well, and clearly identifies the correct model. It succeeds because it searches the entire model space, in a non-stepwise fashion. The prior for δ serves to focus attention on certain elements of the model space, while not totally excluding others. This

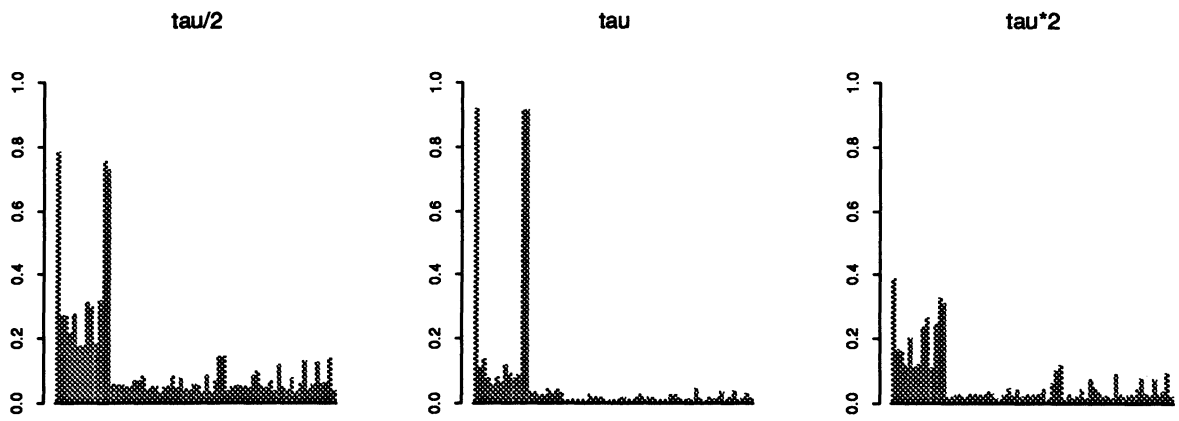

Figure 1: Screening experiment marginal posterior probabilities. The large bars are for factors A, AB, and AC.

class of likely models is much larger than the one used by Hamada and Wu (1992).

4.2 Blood Glucose Experiment

Recall that the blood glucose experiment (also analyzed in Hamada and Wu 1992) consists of continuous factors in either two (factor A) or three levels. The three-level factors are quantitative which allow entertainment of polynomial terms and interactions. In this case there are more variables to consider, thus a more challenging problem. There are 15 degrees of freedom for the main effects, and an additional 98 two-factors interactions to be considered. None are totally correlated, so that all 113 candidate variables will be considered simultaneously. Thus, linear and quadratic terms will be used throughout, with interactions having four components: linear by linear, linear by quadratic, quadratic by linear, and quadratic by quadratic. Choice of ν, λ is based on the automatic procedure of Section 3.2. The hierarchical priors used are:

$$\Pr(\delta_A = 1) = 0.25$$

$$\Pr(\delta_{A^2} = 1|\delta_A) = \begin{cases} 0.01 & \text{if } \delta_A = 0 \\ 0.25 & \text{if } \delta_A = 1 \end{cases}$$

Table 6: Blood glucose experiment posterior model probabilities, relaxed weak heredity prior

model	prob	R^2
BH^2, B^2H^2	0.183	0.7696
B, BH^2, B^2H^2	0.080	0.8548
B, BH, BH^2, B^2H^2	0.015	0.8601
F, BH^2, B^2H^2	0.014	0.7943
GE, BH^2, B^2H^2	0.013	0.8771
AH^2, BH^2, B^2H^2	0.009	0.8528
G^2D, BH^2, B^2H^2	0.009	0.8517
A, BH^2, B^2H^2	0.008	0.7938
B, B^2, BH^2, B^2H^2	0.008	0.8864
H, BH^2, B^2H^2	0.008	0.7855

$$\Pr(\delta_{AB^2} = 1 | \delta_{AB}, \delta_{B^2}) = \begin{cases} 0.01 & \text{if } \delta_{AB} = \delta_{B^2} = 0 \\ 0.10 & \text{if } \delta_{AB} \neq \delta_{B^2} \\ 0.25 & \text{if } \delta_{AB} = \delta_{B^2} = 1 \end{cases}$$

This is a challenging problem, since there are only 18 observations and 113 terms from which to choose. Since there are so many variables, there will likely be many models that fit the data well, and probably quite a few parsimonious ones. The hierarchical priors will be useful here because they will focus attention on good models that also make sense.

The complexity of the problem is apparent in the simulation, which takes much longer to mix sufficiently (i.e., 50,000 runs). When relaxed weak heredity priors are used, the most probable model contains two terms – BH^2 and B^2H^2. The ten most probable models are given in Table 6. This model clearly violates even weak heredity, so there must be a good reason for its large mass. Re-running the algorithm with strict weak heredity (0 replacing probabilities of 0.01) gives the results in Table 7. We see that the best model is a superset of the previous best, with the appropriate terms for weak heredity added (namely B and BH). Other models involving EF also appear possible but less likely.

The fit of this model, indicated by an R^2 of 0.86 is quite good. In fact, both models fit the data better than the R^2 of 0.68 Hamada and Wu (1992) reported for the model with E^2, F^2 and EF, a model that does not obey the current definition of weak heredity. The additional information gained from comparing posteriors originating from weak and strict forms of the prior tells us that it is the higher order interactions between B and H that really drive the model, and may indicate that caution be exercised in identifying a single "best" model.

Table 7: Blood glucose experiment posterior model probabilities, strict weak heredity prior

model	prob	R^2
B, BH, BH^2, B^2H^2	0.146	0.8601
$B, BH, B^2H, BH^2, B^2H^2$	0.034	0.8828
H, H^2, BH^2, B^2H^2	0.033	0.7903
H, BH, BH^2, B^2H^2	0.031	0.7908
F, F^2, DF, D^2F, EF	0.024	0.8835
$H, H^2, AH^2, BH^2, B^2H^2$	0.017	0.8735
B, B^2, BH, BH^2, B^2H^2	0.013	0.8917
B, H, BH, BH^2, B^2H^2	0.013	0.8760
B, H, H^2, BH^2, B^2H^2	0.008	0.8756
E, E^2, CE, EF	0.008	0.6979

One of these models was also identified by Jan and Wang (1994). They identified a model with B, BH^2 and B^2H^2 terms, similar to the one identified by our procedure. In fact the same model can be identified using a modified strategy recommended by Hamada and Wu (1992, p. 136). Both the Hamada-Wu and Jan-Wang procedures find only one optimal model, rather than a set of plausible ones.

4.3 Hard-to-Control Factors Experiment

As discussed in Section 2.3, not all the predictors could be controlled during the experiment, resulting in complex aliasing. In this experiment, the variables 1-7 and 8-11 are considered to be non-interacting groups, so that interactions between them are not entertained. All other interactions are considered. While all the predictors are continuous, only a single quadratic term may be considered because of the structure of the center run (the -1, $+1$ levels all map to 1, resulting in very high correlations between all quadratic terms).

The prior parameters (ν, λ) are chosen to be (2,0.00458), giving a mean of 0.12 for σ, and an upper 99% quantile of 0.66, slightly larger than the unadjusted standard deviation of Y, calculated to be 0.60. The ranges of X_j's will not be used to determine ΔX_j here because of the large outliers in some of the uncontrolled predictors. Instead we shall assume that the original settings of $(-1,+1)$ represent large changes, and take $\Delta X_j = 2$ for all main effects and interactions except X_1^2, which will have a value of 1. Then the $\Delta Y/3\Delta X_i$ rule is used as a starting point for τ. It turns out that for this example, the posteriors have an average of ten active terms with this automatic choice of

Table 8: Hard-to-control factors experiment posterior model probabilities

model	prob	R^2
x1 x3 x1x6 x3x6	0.0425	0.7973
x1 x3 x4 x1x6 x3x6	0.0194	0.8488
x1 x3 x4 x10 x1x6 x3x6	0.0133	0.9034
x1 x3 x6 x1x6 x3x6	0.0114	0.7986
x1 x3 x10 x1x6 x3x6	0.0102	0.8425
x1 x3 x1x3 x1x6 x3x6	0.0081	0.8209
x1 x2 x3 x4 x10 x47	0.0065	0.8886
x1 x3 x4 x10 x1x2	0.0056	0.8206
x1 x3 x4 x7 x10 x2x4	0.0053	0.8854
x1 x3 x4 x6 x1x6 x3x6	0.0052	0.8518

τ. Since more parsimonious models are desired, the procedure was re-run after doubling the value of τ.

Posterior probabilities for the model are displayed in Table 8, and marginal posteriors are given in Figure 2. Although no model has a decisive amount of mass, it appears that the terms 1, 3, 1 x 6, 3 x 6 are important, since they appear in almost all of the most probable models, and have high marginal probabilities of being active.

4.4 Supersaturated Design Experiment

When analyzing the 14-run supersaturated design, we restricted our attention to main effect only models because there are already more effects than observations. Because the supersaturated design is a screening design, this seems to be a reasonable approach. The automatic procedures were used to choose the regression coefficient priors and the hierarchical priors of Section 4.1 were used in the variable selection. The values used were $\lambda = 67, \nu = 2, \tau = 2.31, c = 10$, which appear to produce reasonable results with no modifications. The joint posteriors are given in Table 9.

The results suggest that factors 4, 12, 15, 20 (and perhaps 10) are active, as Lin (1993) found from his analysis. Although the proposed Bayesian methodology did not find any different results, its more thorough search and flexible priors will work to its advantage in larger experiments which have more candidate models.

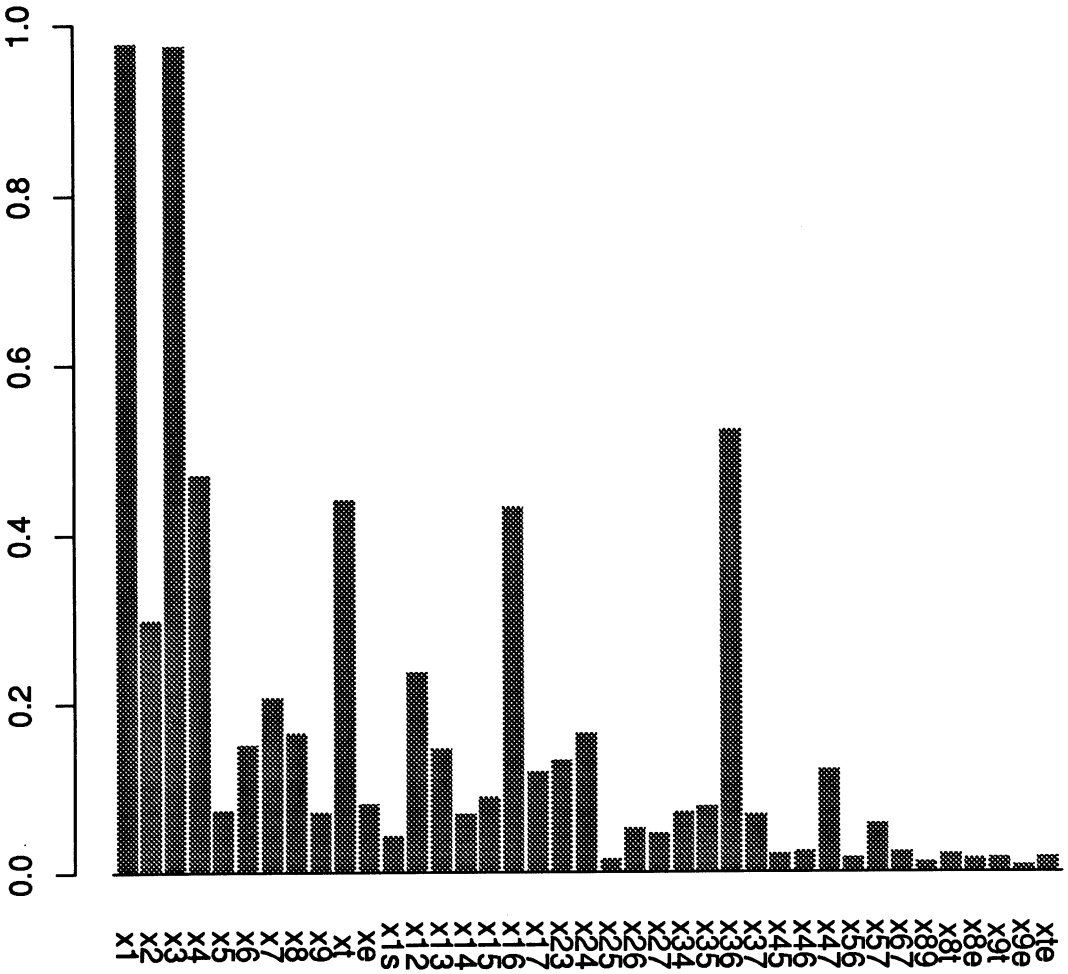

Figure 2: Hard-to-control factors experiment marginal posterior probabilities

Table 9: Supersaturated design experiment posterior model probabilities

Model	prob	R^2
4 12 15 20	0.0390	0.9548
4 10 12 15 20	0.0295	0.9730
4 10 11 12 15 20	0.0195	0.9867
4 12 15 20 21	0.0175	0.9688
4 11 12 15 20	0.0110	0.9661
1 4 12 15 20	0.0105	0.9697
4 10 12 15 20 21	0.0095	0.9826
4 12 13 15 20	0.0080	0.9640
4 7 10 11 12 15 20	0.0070	0.9982
4 12 15 17 20	0.0070	0.9578

5 Discussion

Data with complex aliasing arise in numerous situations: experiments using two-level nongeometric screening designs, multi-level and mixed-level fractional factorial designs and nearly orthogonal designs, experiments with hard-to-control factors and with mistakes in setting the factor levels and experiments using supersaturated designs. Moreover, observational data will typically have complex aliasing. Because data with complex aliasing arise often, an efficient analysis methodology is desirable. Hamada and Wu (1992) showed that information about interactions could be obtained in such situations and proposed an iterative guided stepwise regression strategy.

The current paper presents a more efficient methodology because it searches the model space more thoroughly, much like an all-subsets regression, except that a plausible class of models (i.e., avoiding models with only interactions) are considered. Moreover, the proposed methodology requires much less computation because the search is done stochastically rather than fitting all possible models. A Bayesian approach combined with the recent advances in Bayesian computing provides a quick and easy implementation of this strategy. The flexible hierarchical priors of Section 3.2 provide a powerful way to concentrate the search on a reasonable class of models. Note that previous work on Bayesian model selection (Mitchell and Beauchamp 1988) could not handle interactions. An advantage of the Bayesian approach is that models that the data strongly suggest but fall outside the reasonable class of models defined by the hierarchical priors can be identified. Moreover, the proposed methodology can identify several (perhaps incompatible) models that

explain the data equally well. We used the proposed methodology on Hamada and Wu's (1992) constructed Example 5, a 12-run PB design with true model $Y = 2A + 4C + 2BC - 2CD + \epsilon$, where ϵ is normally distributed with mean zero and standard deviation 0.5. The proposed methodology quickly found several incompatible but equally plausible models (i.e., that obey effect heredity). This illustrates a limitation of the PB designs which may require additional experimentation to resolve. Nevertheless, it is important that the experimenter know that there are several models that fit the data well. Finally, the posterior probability of a model provides a calibrated measure of a model's goodness, but does not require adjustment for the number of effects in the model. In fact, the posterior probability provides better discrimination of various models as seen in Tables 6-9. Note that the best model in the examples not only had a larger posterior probability but interestingly enough, also had a smaller number of effects than the next best model.

Our paper adopted the SSVS algorithm of George and McCulloch (1993) to obtain the model posterior probabilities. Other algorithms such as the Markov Chain Monte Carlo model composition (MC^3) of Raftery, Madigan and Hoeting (1993) could be used in combination with the hierarchical priors of Section 3.2. We feel that the choice of algorithm will not influence the conclusions reached, since both approaches are quite similar in spirit. Both adopt a Bayesian stochastic search algorithm that calculates posterior probability of models based on the samples produced by the algorithm. The main difference is that SSVS defines a "small" effect as near zero, while MC^3defines it as being exactly zero. However, the two crucial elements of our approach may be used with either algorithm: a stochastic search, and a prior on the model space that relates interactions to main effects.

Box and Meyer (1993) proposed an an alternative Bayesian approach focused on factors rather than specific effects. Their proposal can be summarized as follows. Suppose there are k factors. Using an independence prior on the "factors", each factor has prior probability π of being active. Then for each of the 2^k subsets of the k factors, a posterior model probability given the data (for a specific model) is calculated. Say, there are i factors in a particular subset. Then the corresponding model has all the main effects, two-factor and three-factor interactions (provided that i is at least two and three, respectively). Note that the number of effects in some of the models will exceed the number of observations. Box and Meyer (1993) used an independence prior for all the effects, i.e., regression coefficients β, which does not differentiate between main

effects and interactions. The posterior model probabilities are calculated directly which can be computationally intensive. Active factors are identified using marginal posterior probabilities, i.e., the sum of posterior probabilities for all the models given above containing a particular factor. In contrast, our proposed methodology focuses on effects rather than factors, and in addition to marginal posteriors, considers joint posterior probabilities, namely posterior probabilities of models. Our methodology requires less computation than an all-subsets approach or exhaustive search (such as that used by Box and Meyer) because the search through the model space is done stochastically. Moreover, the search is focused on, though not restricted to, a class of reasonable models through the specification of flexible hierarchical priors.

Acknowledgments

H. Chipman was supported by funding from the University of Chicago Graduate School of Business. M. Hamada was supported by research grants from General Motors of Canada Limited, the Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada. C. F. J. Wu was supported by the National Science Foundation DMS 9404300.

References

Box, G. E. P. and Meyer, R. D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.

Box, G. E. P. and Meyer, R. D. (1993), "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, 25, 94–105.

Chipman, H. (1994), "Bayesian Variable Selection with Related Predictors," technical report STAT-94-13, Department of Statistics and Actuarial Science, University of Waterloo. Submitted for publication.

George, E. I. and McCulloch, R. E. (1993), "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

Hamada, M. and Wu, C. F. J. (1992), "Analysis of Designed Experiments with Complex Aliasing," *Journal of Quality Technology*, 24, 130–137.

Hall, M. Jr. (1961), "Hadamard Matrices of Order 16," *Research Summary*, No. 36-10, 1, 21–26, Jet Propulsion Laboratory, Pasadena, CA.

Henkin, E. (1986), "The Reduction of Variability of Blood Glucose Levels" in *Fourth Supplier Symposium on Taguchi Methods*, American Supplier Institute, Inc., Dearborn, MI, 758–785.

Hunter, G.B., Hodi, F.S. and Eager, T.W. (1982), "High-Cycle Fatigue of Weld Repaired Cast Ti-6Al-4V," *Metallurgical Transactions*, 13A, 1589–1594.

Jan, H. W., and Wang, P. C. (1994), "Analysis of Experimental Data from Orthogonal Main-Effect Plans," unpublished manuscript.

Lin, D. K. J. (1993), "A New Class of Supersaturated Designs," *Technometrics*, 35, 28–31.

Mitchell, T. J. and Beauchamp, J. J. (1988), "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83, 1023–1036.

Plackett, R. L. and Burman, J. P. (1946), " The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305–325.

Raftery, A., Madigan, D., and Hoeting, J. (1993), "Model Selection and Accounting for Model Uncertainly in Linear Regression Models," Technical Report No. 262, Department of Statistics, University of Washington.

Smith, A. F. M. and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, 55, 3–23.

Wang, J. C. and Wu, C. F. J. (1991), "An Approach to the Construction of Asymmetrical Orthogonal Arrays," *Journal of the American Statistical Association*, 86, 450–456.

Wang, J. C. and Wu, C. F. J. (1992), "Nearly Orthogonal Arrays with Mixed Levels and Small Runs," *Technometrics*, 34, 409–422.

Williams, K. R. (1968), "Designed Experiments," *Rubber Age*, 100, 65–71.

Wu, C. F. J. (1993), "Construction of Supersaturated Designs Through Partially Aliased Interactions," *Biometrika*, 80, 661–669.