Fast Model Search for Designed Experiments with Complex Aliasing

Hugh A. Chipman

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, N2L 3G1, Canada
hachipma@uwaterloo.ca

Abstract:

In screening experiments, run size considerations can necessitate the use of designs with complex aliasing patterns. Such designs provide an opportunity to examine interactions and other higher order terms as possible predictors, as Hamada and Wu (1992) propose. The large number of model terms mean that many models may describe the data well. Bayesian stochastic search methods that incorporate preferences for certain model structures (Chipman, Hamada, and Wu, 1997) are one effective method for identifying a variety of models. This paper introduces priors that allow the posterior to be simplified, speeding up the search and eliminating Monte Carlo error in the evaluation of model probabilities. Default choices of prior parameters are proposed. Several new plots and summaries are illustrated using simulated data in a Plackett-Burman 12-run layout.

Keywords and phrases: interactions, partial aliasing, Plackett-Burman designs, Gibbs sampler, automatic prior selection.

1 Introduction

When screening many variables, designs with economical run sizes are desirable. Quite often these designs will not be orthogonal if terms other than main effects are considered. A common example is the 12-run Plackett-Burman (1946) (PB) design (see Table 1), in which each main effect is correlated (partially aliased) with the 45 two-factor interactions not involving that effect.

Hamada and Wu (1992) viewed designs with complex aliasing as an opportunity to identify promising interactions as well as main effects. With a modified stepwise algorithm, they showed that promising models could be identified when only a few effects were large (the assumption of *effect sparsity*), and attention was focused on interactions between large main effects (the assumption of *effect heredity*). While their stepwise algorithm finds promising models, it is not always able to identify all promising models. Chipman, Hamada, and Wu (1997)

A	B	C	D	E	F	G	H	I	J	K	Y_1	Y_2
+	+	−	+	+	+	−	−	−	+	−	1.058	-1.358
+	−	+	+	+	−	−	−	+	−	+	1.004	1.228
−	+	+	+	−	−	−	+	−	+	+	-5.200	2.291
+	+	+	−	−	−	+	−	+	+	−	5.320	9.432
+	+	−	−	−	+	−	+	+	−	+	1.022	-5.719
+	−	−	−	+	−	+	+	−	+	+	-2.471	-2.417
−	−	−	+	−	+	+	−	+	+	+	2.809	-2.494
−	−	+	−	+	+	−	+	+	+	−	-1.272	2.674
−	+	−	+	+	−	+	+	+	−	−	-0.955	-5.943
+	−	+	+	−	+	+	+	−	−	−	0.644	1.596
−	+	+	−	+	+	+	−	−	−	+	-5.025	6.682
−	−	−	−	−	−	−	−	−	−	−	3.060	-5.973

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

(CHW) gave a more thorough model search algorithm. This approach used a Bayesian model, and the Stochastic Search Variable Selection (SSVS) algorithm of George and McCulloch (1993) (henceforth GM93). This methodology was able to search the model space more completely while incorporating effect heredity.

This paper considers a number of enhancements to CHW. First, it uses a "fast" version of SSVS proposed by George and McCulloch (1997) (henceforth GM97). This formulation allows analytical evaluation of relative probabilities on models, rather than relying upon Monte Carlo integration, as in CHW. The ability to analytically evaluate relative probabilities also helps identify stopping conditions for the algorithm. Second, the priors used are discussed in more detail, and automatic choices for all prior parameters are given, simplifying the use of this approach for non-Bayesians who wish to identify promising models. Third, a number of graphics and summaries are introduced to identify promising models found by the procedure. These include a C_p - like plot that divides up model probability according to model size, a time series plot of the probability of models visited, and reweighting of the posterior to assess prior influence.

The paper is organized as follows. Section 2 outlines the priors used, and in Section 3 the Gibbs sampler is described for this situation. An example in Section 4 illustrates automatic choices of prior parameter. Section **??** gives an example in which multiple models fit the data well. The impact of different priors on models is examined, and several new plots and summaries are introduced to examine the promising models identified. Section **??** discusses related issues.

2 Priors

For the regression model, $Y = X\beta + \sigma\epsilon$, with ϵ a standard normal random variable, the parameters of interest are β, σ. Some of the k columns of X (such as interactions) may be formed from the original variables.

In variable selection, some of the k elements of β may be negligible. This is captured in GM93 by augmenting the model with an unobserved k-vector γ of 0's and 1's. A 0 corresponds to a negligible coefficient, and a 1 corresponds to an important coefficient. The prior is specified as

$$\Pr(\gamma, \sigma, \beta) = \Pr(\beta|\sigma, \gamma) \Pr(\gamma) \Pr(\sigma)$$

with

$$\beta_i \sim \text{ i.i.d. } N(0, \sigma^2\tau_i^2(1-\gamma_i) + \sigma^2\tau_i^2 c_i^2 \gamma_i), \qquad i = 1, \ldots, k$$

By choosing $c_i \gg 1$, the prior variance of β_i is $\sigma^2\tau_i^2$ when $\gamma_i = 0$ and c_i^2 times larger when $\gamma_i = 1$. The hyperparameters c_i, τ_i are chosen to indicate magnitudes of small and large effects. Section 4 suggests how these may be selected automatically.

As in GM93,

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

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

This prior formulation differs from that of GM93 and CHW, in that the variance of β_i depends on σ. With this prior, the posterior for (β, σ, γ) can be integrated over σ, β. As discussed in Section 3, this facilitates model search. The posterior for γ is given by (see GM97):

$$\Pr(\gamma|Y) \propto |\tilde{X}'\tilde{X}|^{-1/2} |D_\gamma|^{-1} (\lambda\nu + S_\gamma^2)^{(n+\nu)/2} \Pr(\gamma) = g(\gamma), \tag{1}$$

where D_γ is diagonal with ith element $\tau_i(1-\gamma_i) + c_i\tau_i\gamma_i$,

$$S_\gamma^2 = \tilde{Y}'\tilde{Y} - \tilde{Y}'\tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y},$$

$$\tilde{Y} = \begin{bmatrix} Y \\ 0 \end{bmatrix} \qquad \text{and} \qquad \tilde{X} = \begin{bmatrix} X \\ D_\gamma^{-1} \end{bmatrix}.$$

The remaining component of the prior is for γ, which puts probability on the space of all possible models. When interactions and other related predictors are present, it is not practical to assume that all elements of γ are independent. Instead, a dependence structure for related predictors (Chipman (1996)) is used. This prior consists of a product of k probabilities:

$$\Pr(\gamma) = \prod_{i=1}^{k} \Pr(\gamma_i|\text{Parents}(\gamma_i)) \tag{2}$$

The probability that a given term is active or inactive depends on its "parent" terms, typically taken to be those terms of the next lowest order from which the given term may be formed. For example, main effects A and B would have no parents, and an interaction AB would have two parents (A and B). The corresponding elements of (2) would be $\Pr(\gamma_A), \Pr(\gamma_B)$ and $\Pr(\gamma_{AB}|\gamma_A, \gamma_B)$. The prior is specified by choosing marginal probabilities that a main effect is active, and the conditional probability that an interaction is active:

$$P(\gamma_{AB} = 1|\gamma_A, \gamma_B) = \begin{cases} p_{00} & \text{if } (\gamma_A, \gamma_B) = (0,0) \\ p_{01} & \text{if } (\gamma_A, \gamma_B) = (0,1) \\ p_{10} & \text{if } (\gamma_A, \gamma_B) = (1,0) \\ p_{11} & \text{if } (\gamma_A, \gamma_B) = (1,1) \end{cases}. \tag{3}$$

Choosing $p_{00} = p_{10} = p_{01} = 0, p_{11} > 0$ allows an interaction to be active only if both corresponding main effects are active (referred to as *strong heredity*). Choosing $p_{00} = 0, p_{01}, p_{10}, p_{11} > 0$ allows an interaction to be active if one or more of its parents are active (*weak heredity*). Models obeying strong heredity are usually easier to interpret, while weak heredity may help the stochastic search explore the model space. Typical values of $(p_{00}, p_{10}, p_{01}, p_{11})$ might be (0, 0, 0, 0.25) for strong heredity, and (0, 0.10, 0.10, 0.25) for weak heredity. Values less than 0.5 represent the belief that only a few effects are likely to be active (see Box and Meyer (1986) for a similar argument).

3 Efficient Stochastic Search

The large number of models makes it impractical to exhaustively evaluate the posterior probability for each model. Instead the Gibbs sampler is used to make draws from the posterior distribution. In the context of model selection, this algorithm may be thought of as a stochastic search.

For this problem, the Gibbs sampler (see Smith and Roberts (1993) and Gelfand and Smith (1990)) may be summarized as follows. The algorithm makes use of "full conditional" distributions, which specify the conditional distribution of one parameter given all others (and the data). The algorithm starts with initial values of all parameters, and then repeatedly draws each parameter conditional on all the others and the data:

0. Start with $\gamma^0 = (\gamma_1^0, \gamma_2^0, \ldots, \gamma_k^0)$

1. Draw γ_1^1 from $p(\gamma_1|\gamma_2^0, \ldots, \gamma_k^0, Y)$.

2. Draw γ_2^1 from $p(\gamma_2|\gamma_1^1, \gamma_3^0, \ldots, \gamma_k^0, Y)$

   $\vdots$

k. Draw γ_k^1 from $p(\gamma_k|\gamma_1^1, \ldots, \gamma_{k-1}^1, Y)$.

Each draw is from a Bernoulli distribution. Steps 1 to k are repeated a large number of times, each time conditioning on the most recently drawn values of the other elements of γ. The sequence $\gamma^0, \gamma^1, \gamma^2, \ldots$ converges to the posterior for γ.

A similar scheme was used in CHW, but with the parameters β and σ included in the Gibbs sampler rather than being integrated out. Elimination of β and σ from the Gibbs sampler speeds up the algorithm (see GM97 and Liu, Wong, and Kong (1995)). By reducing the dimensionality of the parameter space, the Gibbs sampler is able to move around faster.

An additional advantage is the ability to analytically evaluate the posterior probability of a model up to a normalizing constant, using (1). The posterior probability of a model γ' is then

$$\Pr(\gamma'|Y) = g(\gamma') \Bigg/ \sum_{i=1}^{2^k} g(\gamma_i). \tag{4}$$

The normalizing constant in the denominator entails evaluation of the posterior probability for all models, which is prohibitive. For any given set of models, relative probabilities can be evaluated by replacing the denominator of (4) with a sum of g over all models visited so far. Probabilities estimated in this way will be too large, but if most of the "good" models have been visited, this provides a way to remove sometimes sizeable Monte Carlo error. CHW experienced this error because their posterior probability estimates were based on frequencies.

4 Automatic selection of prior parameters

In this section, an example illustrates automatic choices for prior parameters $c_i, \tau_i, \nu, \lambda$, and the model prior for γ. The choices proposed here are easier to implement than those described in CHW.

The data are simulated from a simple model considered by Hamada and Wu (1992): $Y = A + 2AB + 2AC + \sigma\epsilon$ with $\sigma = 0.25$ and ϵ a standard normal. Eleven predictors are arranged in a 12-run PB design, shown in Table 1 along with the response Y_1.

In choosing the prior for σ^2, ν and λ are chosen so the upper tail is some large value and the middle is near the anticipated residual variance. The prior expected value of σ^2 is

$$\mathrm{E}(\sigma^2) = \frac{\lambda\nu}{\nu - 2} \qquad \text{for } \nu > 2,$$

suggesting that λ be chosen near the expected residual variance. In the absence of expert knowledge, some fraction of the unconditional variance of the response could be used to estimate λ. CHW propose

$$\boxed{\lambda = \mathrm{Var}(Y)/25.}$$

ν	mean	0.01	0.1	0.5	0.9	0.99
1	–	0.15	0.37	2.2	63.33	6365
2	–	0.22	0.43	1.44	9.49	99.50
3	3	0.26	0.48	1.27	5.13	26.13
4	2	0.30	0.51	1.19	3.76	13.46
5	1.67	0.33	0.54	1.15	3.10	9.02
6	1.5	0.36	0.56	1.12	2.72	6.88
7	1.4	0.38	0.58	1.10	2.47	5.65
8	1.33	0.40	0.60	1.09	2.29	4.86
9	1.29	0.42	0.61	1.08	2.16	4.31
10	1.25	0.43	0.63	1.07	2.06	3.91

Table 2: Quantiles of an Inverse Gamma distribution with $\lambda = 1$

In this example, $\mathrm{Var}(Y) = 10$ so $\lambda = 0.40$ is selected.

The parameter ν acts as degrees of freedom, with larger values corresponding to a distribution that is tighter about λ. A sufficiently diffuse prior may be selected by choosing ν so that the upper tail (say the 99th percentile) is roughly equal to the unconditional variance. Table 2 gives various quantiles for an inverse gamma with $\lambda = 1$. Choosing $\nu = 5$ would place the 99th percentile of the prior at 9.02λ, for example. Smaller degrees of freedom are possible (for example values of $\nu = 1.5$ were used in CHW), although they can lead to unreasonably long tails, because $\mathrm{Var}(\sigma^2)$ is not defined for $\nu \leq 4$. In general, one would choose

$$\boxed{\nu = 5 \quad \text{or from Table 2.}}$$

The prior parameters c_i and τ_i may be chosen as

$$\boxed{c_i = 10, \tau_i = \tfrac{1}{3 \times \mathrm{range}(X_i)}.} \tag{5}$$

Box and Meyer (1986) suggest $c = 10$, separating large and small coefficients by an order of magnitude. Choice of τ_i is motivated by the fact that a small coefficient has standard deviation $\sigma\tau_i$, and will lie within $0 \pm 3\sigma\tau_i = 0 \pm \sigma/\mathrm{range}(X_i)$ with very high probability. Even a large change in X_i (say of magnitude comparable to $\mathrm{range}(X_i)$) will result in a change in Y of no more than σ, which is presumably small.

The variable selection procedure can be quite sensitive to the choice of τ_i (see CHW and GM93). Equation (5) captures the relative magnitudes of the τ_i for different variables, but the overall magnitude may need tuning. Box and Meyer (1993) and CHW propose methods for tuning based on runs of the search algorithm. A faster alternative based on predictive distributions is proposed here. For any given model γ, the expected value of Y for a given X may be

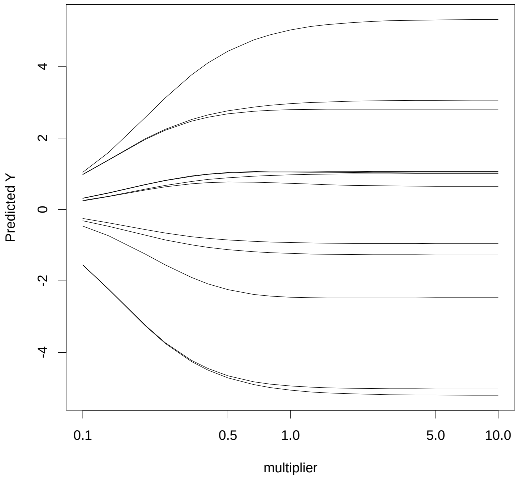

Figure 1: Predicted response under original design for various multiples of parameter τ

calculated. The magnitude of τ_i will determine the degree of shrinkage for coefficient β_i, in a manner similar to ridge regression. A simple way to assess the value of τ_i is to see how the predictions vary for a range of values $r\tau_i$ for a single given model. A good τ_i value would be the smallest value not shrinking predictions too much.

The posterior mean for β is given by

$$\hat{\beta}_\gamma = (X'X + D_\gamma^{-2})^{-1}X'Y$$

where D_γ is diagonal with elements $\tau_i(1-\gamma_i) + \tau_i c_i \gamma_i$. See GM97 for details.

Figure **??** plots predicted values for the original 12 design points for $r \in (1/10, 10)$. The model used (A, B, C, AB, AC) was identified by stepwise regression and a subsequent "hereditization". The "1" value on the horizontal axis is the default choice (5) for τ. In this case, the default seems quite reasonable, as any smaller multiples would shrink too much.

The prior on γ may be selected by considering the prior expected number of active terms. If we have

$k =$ number of main effects
$p = \Pr(\text{main effect active})$
$p_{01} = p_{10} = \Pr(\text{interaction active}|\text{one of its parents is active})$
$p_{11} = \Pr(\text{interaction active}|\text{both its parents are active})$,

then under weak heredity

$$\begin{aligned} \mathrm{E}(\#\text{ effects}) &= \mathrm{E}_f[\mathrm{E}(\#\text{ effects}|f \text{ main effects active})] \\ &= \mathrm{E}_f(f + \binom{f}{2} p_{11} + f(k-f)p_{01}) \end{aligned} \quad (6)$$

The first element is the number of main effects. The second is the expected number of active interactions with both parents active. There are $\binom{f}{2} = f(f-1)/2$ such terms and each one has a probability of p_{11} of being active. The third term is the expected number of active interactions with exactly one parent active. These are interactions between one of f active parents and $k-f$ inactive parents. Consequently there are $f(k-f)$ allowable terms, and each has probability p_{01} of being active. Since f is Binomial(k,p), $\mathrm{E}(f) = kp$ and $\mathrm{E}(f^2) = kp(1-p+kp)$. Simplification of (**??**) gives:

$$\begin{aligned} \mathrm{E}(\#\text{ effects}) &= kp + p_{11}\mathrm{E}(f(f-1)/2) + p_{01}\mathrm{E}(fk - f^2) \\ &= kp + p_{11}k(k-1)p^2/2 + p_{01}k(k-1)p(1-p) \end{aligned} \quad (7)$$

In this example strong heredity is used so 2.75 active main effects and 0.86 active interactions are expected.

1000 draws from the posterior on γ are generated using the Gibbs sampler and the integrated posterior. A total of 333 different models were visited. The most probable model is A, B, C, AB, AC (probability = 0.55), followed by A, B, C, AB, AC, BC (0.06) and eight other models, each of which contains A, B, C, AB, AC, and one other main effect (roughly 0.03 probability each). It is quite clear that A, B, C, AB, AC are active. Although main effects B and C are not important, they are included by the strong heredity prior.

5 A more difficult problem

The design for this example remains the same, and the new response Y_2 is given in Table 1. Data are generated from $Y = 2A + 4C + 2BC + 2CD + \sigma\epsilon$ with $\sigma = 0.5$ and ϵ a standard normal.

Prior parameters $\nu = 5, \lambda = \mathrm{Var}(Y)/25, c_i = 10, \tau_i = 1/3\mathrm{range}(X_i)$ were chosen as in the previous section. A number of priors on the model space are explored:

1. Strong heredity with $(p, p_{11}, p_{01}) = (.25, .25, 0)$, yielding 3.61 effects expected to be active.

2. Weak heredity with $(p, p_{11}, p_{01}) = (.25, .25, .10)$, yielding 5.67 effects expected to be active.

3. Independence prior with $(p, p_{11}, p_{01}) = (0.25, 0.1, 0.1)$ and $p_{00} = 0.10$. This yields $2.75 + 55 \times (0.10) = 8.25$ effects expected to be active.

Strong Heredity

Model	Prob	R^2
C	0.183	0.712
C J	0.077	0.794
C H	0.059	0.782
C G	0.045	0.77
C H J	0.027	0.864
C D C:D	0.024	0.84
C G J	0.019	0.852
B C H B:H	0.033	0.923
C G J C:G	0.019	0.928
C H J C:H	0.012	0.914
B C H B:C B:H	0.017	0.958
C D G C:D D:G	0.01	0.948
B C H J B:H	0.006	0.957
B C H I B:C B:H	0.003	0.989
B C H B:C B:H C:H	0.002	0.961
C D G C:D C:G D:G	0.002	0.961

Weak Heredity

Model	Prob	R^2
C	0.221	0.712
C C:D	0.065	0.812
C J	0.04	0.794
C B:C	0.031	0.782
C H B:H	0.028	0.887
C E E:I	0.021	0.917
C I E:I	0.021	0.917
A C B:C C:D	0.02	0.992
C D C:D D:G	0.012	0.927
B C B:C B:H	0.01	0.923
A C D B:C C:D	0.002	0.994
C G H B:H G:H	0.002	0.953
B C B:C B:H C:F	0.002	0.961

Table 3: Models identified under strong and weak heredity priors. The most probable models of each size are given.

1000 iterations of the Gibbs sampler were used in each of the three cases. The best models of each size for each of the three priors are displayed in Tables **??** and **??**. Since a variety of priors on model size seem plausible, this dependence can be reduced by reporting results conditional on model size, rather than just giving the most probable models (as in CHW).

The main effect for C is active, but models containing a variety of other terms explain the data well. Relaxing the prior (from strong to weak or weak to independence) produces models with better fit for a given number of effects. These models may be more difficult to interpret, especially in the independence case where numerous two way interactions are included without any corresponding main effects.

The true model (A, C, BC, CD) is the most likely of its size in the weak heredity case. The true model has no probability in the strong heredity case. The closest model under strong heredity would be A, B, C, D, BC, CD, which was not visited in the first 1000 steps. The size of this model (almost twice the number of active terms as expected) and the many six term models that fit well may explain why it was not visited.

Figure **??** gives marginal probabilities for the weak heredity case. The main effect for C has the largest probability of being active, and other possibly active effects include H, J, BC, BH, CD. The fact that none of the latter effects

Independence		
Model	Prob	R^2
C H	0.068	0.782
C E	0.016	0.712
C J:K	0.013	0.746
C J E:I	0.115	0.935
C B:H E:J	0.101	0.923
C H E:I	0.09	0.929
C H J E:I	0.03	0.953
C H E:I I:K	0.025	0.965
C G J E:I	0.022	0.947
B C J A:F D:F	0.007	0.993
C G H A:B D:G	0.005	0.991
A B C D:G I:J	0.004	0.955
C D J E:K F:I G:H	0.002	0.995
C E H B:E B:H E:J	0.001	0.991
C D E B:H C:F E:J	0.001	0.992

Table 4: Models identified under independence prior. The most probable models of each size are given.

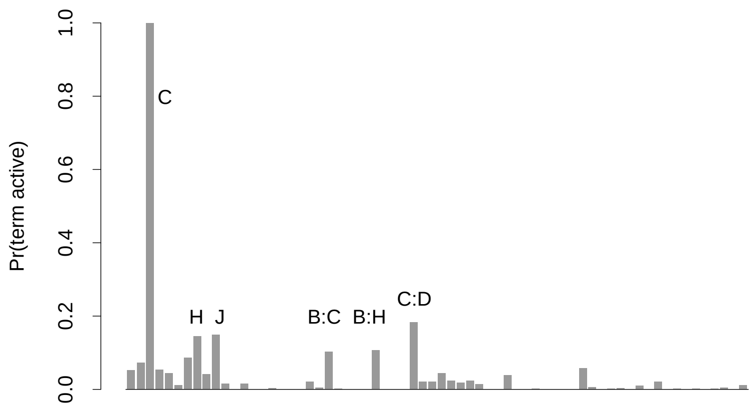

Figure 2: Marginal probability that an effect is large, weak heredity case.

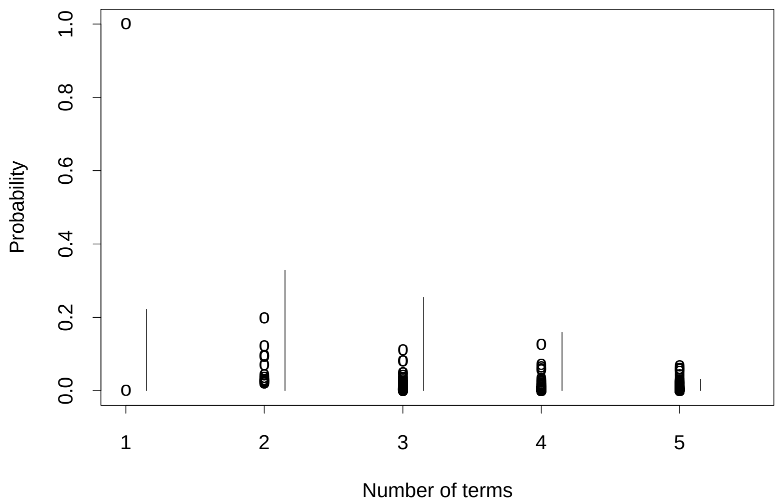

Figure 3: The most probable models ordered by size. The vertical bars give total probability for all models of this size and the dots give probabilities for specific models conditional on that model size.

Model	Prob	Unweighted Prob	R^2
C	0.221	0	0.712
C C:D	0.065	0.001	0.812
C J	0.04	0.001	0.794
C B:C	0.031	0.001	0.782
C H	0.03	0.001	0.782
C H B:H	0.028	0.005	0.887
C G	0.023	0	0.77
C E E:I	0.021	0.004	0.917
C I E:I	0.021	0.004	0.917
A C B:C C:D	0.02	0.031	0.992

Table 5: Posterior probabilities on models, weak heredity case. Second "unweighted" probability column corresponds to a weak heredity prior with all nonzero probabilities equal to 0.5.

has a large probability indicates considerable uncertainty about which effects in addition to C are active. Figure **??** is a C_p-like plot, giving the conditional probability of models, conditional on model size (plotted as circles). The marginal probability associated with all models of a certain size is represented with a vertical line. This plot indicates that the most likely model sizes are 1, 2, 3 or 4 terms. Among the one term models, the model with C only dominates, but in other cases a number of models are close.

The different models identified in Tables **??** and **??** indicate the strong influence of the prior. To illustrate this effect, another weak heredity prior is used with $p = p_{01} = p_{11} = 0.5$. This prior is uniform on all weak heredity models. Rather than re-run the Gibbs sampler, probabilities are recalculated (Table **??**) for models found with the original prior. The large change in the posterior probabilities indicates that the prior penalizes larger models heavily.

One practical issue is how long to run the chain. When probable models are no longer being discovered, the chain can be stopped. Figure **??** presents these relative probabilities for an extended run of 5000 iterations, under strong heredity priors. Each vertical line is the probability of a model the first time it is visited. After the first 1000 iterations, new models are still being visited, but none with appreciable probability, suggesting that 1000 runs is sufficient. The solid line gives the cumulative probability, and indicates that about 75% of the probability thus far identified was discovered in the first 1000 steps.

In many problems the algorithm is quite fast and computing time inexpensive enough that longer runs are practical. In situations where as many models as possible are desired, long runs would be preferred. Each of the 1000 iteration

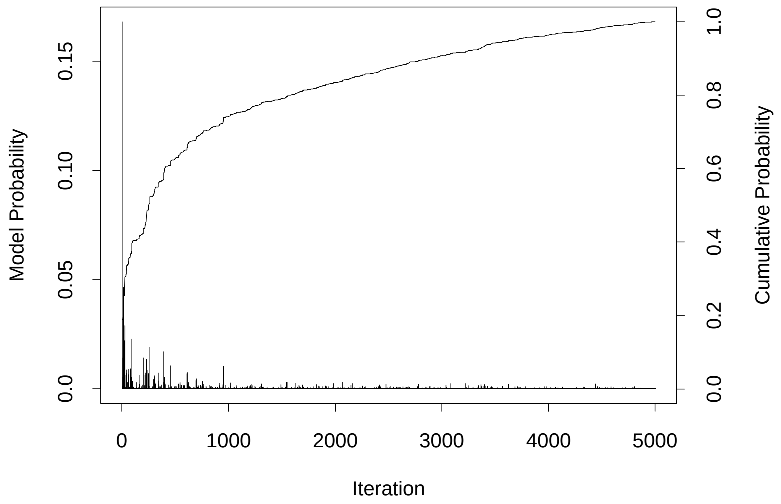

Figure 4: Time series plot of relative probabilities of models on the first time each is visited. The solid line represents the cumulative probability of all models visited so far.

runs here took 50 seconds on a 200 Mhz Pentium Pro system. The relatively small number of iterations used here (1000) was chosen to illustrate that considerable information could be quickly extracted from the data.

6 Discussion

This paper illustrates that choice of prior parameters for variable selection is relatively straightforward. Considerations such as the implied shrinkage of predictions, anticipated number of active terms, and anticipated residual error can be used to provide semi-automatic choices of prior parameters.

Interpretation of output is also an issue, especially when many models may describe the data well. The posterior can sometimes favor models with a certain number of active terms. To explore a wider range of model sizes, graphics and tables that give the best models conditional on size are useful.

Stochastic model search using an integrated posterior offers several advantages over Markov chain Monte Carlo methods in which all parameters are drawn. Convergence will be faster because fewer parameters are sampled at each iteration. The relative log posterior probability of a model is available at any point in the simulation, aiding the decision of when to stop the chain. One disadvantage of this approach is that it will not generalize as easily to other models (generalized linear models, for example). In such cases it may be difficult to integrate the posterior over parameters such as β. Gibbs sampling methods which draw all parameters may be more promising in such situations.

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, 94–104.

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors", *Canadian Journal of Statistics*, 24, 17–36.

Chipman, H., Hamada, M. and Wu, C. F. J., (1997) "A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing", *Technometrics*, 39, 372–381.

Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities",*Journal of the American Statistical Association*, 85, 398–409.

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

George, E. I. and McCulloch. R. E. (1997) "Approaches to Variable Selection", Statistica Sinica, 7, 339–373.

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

Liu, J. S., Wong, W. H. and Kong, A. (1995), "Covariance Structure and Convergence Rate of the Gibbs Sampler with Various Scans", *Journal of the Royal Statistical Society, Series B*, 57, 157–169.

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

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.