# 行政院國家科學委員會專題研究計畫 成果報告

## 在工程實驗上的貝氏模式選取
## 研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 97-2118-M-029-001-
執 行 期 間 ：97 年 08 月 01 日至 98 年 07 月 31 日
執 行 單 位 ：東海大學統計學系


計 畫 主 持 人：俞一唐

計畫參與人員：碩士班研究生-兼任助理人員：卓原鼎
碩士班研究生-兼任助理人員：黃琪嵎
碩士班研究生-兼任助理人員：王德綱
碩士班研究生-兼任助理人員：張哲論
大專生-兼任助理人員：陳柔佐
大專生-兼任助理人員：劉楚頤
大專生-兼任助理人員：鄒振宏


報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 98 年 10 月 29 日

This report is a short version of a working paper entitled "Screening Active Factors by Using Bayesian Computer Experimental Models".

## Introduction

We are concerned with a screening experiment with $p$ candidate factors. When only main effects and 2fi effects are taken consideration, there are $2^{p+\frac{p(p-1)}{2}}$ possible models. It is very common to have say seven to ten factors in an screening experiment (usually more). Then, $2.7 \times 10^9$ and $3.6 \times 10^{16}$, with seven and ten factors respectively, possible models are resulted. If 3fi effects are also considered, then $9.2 \times 10^{18}$ and $4.2 \times 10^{52}$ possible models should be considered ! Although, some empirical principles, such as effect heredity principle (Wu and Hamada, 2000), can rule out most of the models, the number of candidates is still considerable.

The reason why the number of possible models increases dramatically is that the increase of the number of effects. In a screening experiment, what we concern the most is if a factor is active or not. A larger scale follow-up experiment will always be constructed. In the follow-up experiment, the factors selected in the screening experiment are set at more levels, and the different types of effects can be studied more precisely. Hence, in a screening experiment, we should focus on factors instead of effects. In this project, we develop a factor screening method under a Bayesian framework which is using a computer experimental model. Because in the model we do not consider effects, the number of all possible models is much smaller.

## Models

Suppose a screening experiment with $p$ factors and $n$ runs. Each factor can be either active or inner and there are hence $q = 2^p$ possible models, labeled as $M_1, \cdots, M_q$ (including the null model). According to Bayes theorem, the posterior probability of the model $M_k$ given the data $\boldsymbol{y} = (y_1, \cdots, y_n)'$ is given by

$$p(M_k|\boldsymbol{y}) = \frac{p(M_k)f(\boldsymbol{y}|M_k)}{\sum_{i=1}^{q} p(M_i)f(\boldsymbol{y}|M_i)},$$

where $p(M_i)$ is the prior probability of model $M_i$, and $f(\boldsymbol{y}|M_i)$ is the marginal density of $\boldsymbol{y}$ given the model $M_i$.

For a given model $M_i$, let $\boldsymbol{\theta}_i$ denote the associated vector of parameters. Under a Bayesian framework, we select a prior distribution of the model specific parameters $\boldsymbol{\theta}_i$,denoted as $f(\boldsymbol{\theta}_i|M_i)$, and the likelihood function can be denoted as $f(\boldsymbol{y}|M_i, \boldsymbol{\theta}_i)$. Integrating out the parameters, the marginal density of $\boldsymbol{y}$ given the model $M_i$ is expressed as

$$f(\boldsymbol{y}|M_i) = \int_{\Omega_i} f(\boldsymbol{y}|M_i, \boldsymbol{\theta}_i)f(\boldsymbol{\theta}_i|M_i)d\boldsymbol{\theta}_i, \tag{1}$$

where $\Omega_i$ is the set of all possible values of $\theta_i$.

Traditionally, the experimental result is analyzed by using regression, and the responses under various experimental settings are assumed to follow normal distribution independently, which is

$$\boldsymbol{y} \sim N(\boldsymbol{\eta}(\boldsymbol{x}), \frac{1}{\lambda}\boldsymbol{I}),$$

where $\boldsymbol{\eta}(\boldsymbol{x}) = (\eta(\boldsymbol{x}_1), \cdots, \eta(\boldsymbol{x}_n))'$ is the vector of the means under the $n$ experimental settings, $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$, and $\frac{1}{\lambda}$ is the common variance. When various types of effects are considered, $\eta(\boldsymbol{x})$ is modeled as a linear function of various effects (usually main effects and 2fi effects, sometimes 3fi effects) Instead of this, we adopt a Gaussian process (GP) model for modeling $\boldsymbol{\eta}(\boldsymbol{x})$, which is suitable for modeling complex relationships. Such models are commonly employed in geostatistics (Diggle and Ribeiro, 2007) and in computer experiments (Santner et al., 2003). We assume that

$$\eta(\boldsymbol{x}) \sim GP(\mu, \Delta\frac{1}{\lambda}R(\boldsymbol{h})),$$

where $\Delta$ is a turning parameter used to present the signal and noise ratio. The sensitivity to the choice of $\Delta$ will be discussed later. The covariance function is defined as $cov[\eta(\boldsymbol{x}+\boldsymbol{h}), \eta(\boldsymbol{x})] = \Delta\frac{1}{\lambda}R(\boldsymbol{h})$, where $\boldsymbol{h} = (h_1, \cdots, h_p)' \in \mathcal{R}^p$ and $R(\boldsymbol{h})$ is a correlation function. Various types of correlation functions can be found in the literature and we adopt, in this project, the Gaussian product correlation function which is $R(\boldsymbol{h}) = \exp\{-\sum_{l=1}^{p}\gamma_l h_l^2\}$, where $\gamma_l$ is the correlation parameter representing the strength of factor $l$'s impact. Factor levels are usually recoded in the interval $[-1, 1]$. Thus, we reparameterize $\gamma_l = -\frac{1}{4}\log\rho_l$ so that $\rho_l$ represents the correlation coefficient between the responses when the factor $l$ is changed from $-$ to $+$. We denote $\boldsymbol{\theta} = (\mu, \lambda, \rho_1, \cdots, \rho_p)'$,, the collection of all the parameters. Recall that the $2^p$ candidate models are obtained by considering each factor is active or inner. Then, for $M_i$, $\boldsymbol{\theta}_i$ can be obtained by dropping those $\rho_l$'s whose associated factor is inner. Hence,

$$\boldsymbol{y}|M_i, \boldsymbol{\theta}_i \sim N(\mu\mathbf{1}, \frac{1}{\lambda}(\boldsymbol{I} + \Delta\boldsymbol{R}_i)), \tag{2}$$

where $\mathbf{1}$ is the vector of 1's with length $n$ and $\boldsymbol{R}_i$ is the correlation matrix in $M_i$. Letting $\boldsymbol{R}$ be an $n \times n$ matrix with elements $\boldsymbol{R}_{ij} = R(\boldsymbol{x}_i - \boldsymbol{x}_j)$, then $\boldsymbol{R}_i$ can be obtained by replacing $\rho_l = 1$ for those inner factors in $M_i$ in $\boldsymbol{R}$.

For considerations of prior selections, both a model space prior, $p(M_i)$, and model specific parameter priors $f(\boldsymbol{\theta}|M_i)$ need to be specified. For the prior on model space, an empirical principle, named sparsity principle, states that there are only a few active factors in an experiment. Hence, an equally likely prior over all $2^p$ models is not suitable. Letting $f_i$ be the number of active factors in model $M_i$, and $\pi$ be the prior probability that any particular factor is active, we adopt the following prior

$$p(M_i) = \pi^{f_i}(1-\pi)^{p-f_i}.$$

In practice, the number of active factors identified in a screening experiment is less than a half of all the factors considered, so $\pi$ is chosen in the range $(0, 0.5]$. If further information to which factor is more likely to be active, $\pi$ can be chosen dependent on factors. Turning to the choice of parameter priors, let $A_i$ denote the index set of active factors in $M_i$. Given model $M_i$, we specify the following independent and noninformative prior

$$f(\boldsymbol{\theta}_i|M_i) = f(\mu)f(\lambda)\prod_{l\in A_i}\pi(\rho_l), \tag{3}$$

where

$$\begin{aligned} f(\mu) &\propto 1, \\ f(\lambda) &\propto \tfrac{1}{\lambda}, \\ f(\rho_l) &= 1 \text{ as } \rho_l \in (0,1). \end{aligned}$$

According to (1), (2) and (3),

$$
\begin{aligned}
f(\boldsymbol{y}|M_i) &= \int\int\int (2\pi)^{-\frac{n}{2}} |\boldsymbol{I}+\Delta\boldsymbol{R}_i|^{-\frac{1}{2}} \exp\left\{-\frac{\lambda}{2}(\boldsymbol{y}-\mu\mathbf{1})'(\boldsymbol{I}+\Delta\boldsymbol{R}_i)^{-1}(\boldsymbol{y}-\mu\mathbf{1})\right\} d\boldsymbol{\rho} d\lambda d\mu \\
&= CK_i
\end{aligned}
$$

where $C = (2\pi)^{-\frac{n-1}{2}}\Gamma(\frac{n}{2})$ and

$$
K_i = \int |\boldsymbol{I}+\Delta\boldsymbol{R}_i|^{-\frac{1}{2}} \left\{\frac{2\mathbf{1}'(\boldsymbol{I}+\Delta\boldsymbol{R}_i)^{-1}\mathbf{1}}{(\mathbf{1}'(\boldsymbol{I}+\Delta\boldsymbol{R}_i)^{-1}\mathbf{1})(\boldsymbol{y}'(\boldsymbol{I}+\Delta\boldsymbol{R}_i)^{-1}\boldsymbol{y}) - (\mathbf{1}'(\boldsymbol{I}+\Delta\boldsymbol{R}_i)^{-1}\boldsymbol{y})^2}\right\}^{\frac{n}{2}} d\boldsymbol{\rho}. \tag{4}
$$

The integral can be computed by using either numerical integral or Monte Carlo method. A simple approximation of (4) can be implemented by plugging $\boldsymbol{\rho}$ with given values into $\boldsymbol{R}_i$ and obtain $\hat{\boldsymbol{R}}_i$. We are then given

$$
\hat{K}_i = |\boldsymbol{I}+\Delta\hat{\boldsymbol{R}}_i|^{-\frac{1}{2}} \left\{\frac{2\mathbf{1}'(\boldsymbol{I}+\Delta\hat{\boldsymbol{R}}_i)^{-1}\mathbf{1}}{(\mathbf{1}'(\boldsymbol{I}+\Delta\hat{\boldsymbol{R}}_i)^{-1}\mathbf{1})(\boldsymbol{y}'(\boldsymbol{I}+\Delta\hat{\boldsymbol{R}}_i)^{-1}\boldsymbol{y}) - (\mathbf{1}'(\boldsymbol{I}+\Delta\hat{\boldsymbol{R}}_i)^{-1}\boldsymbol{y})^2}\right\}^{\frac{n}{2}}. \tag{5}
$$

Some other more accurate approximation methods of obtaining the marginal density can be found in Kass and Raftery (1995).

Letting $P_j$ be the posterior probability that factor $j$ is active, then

$$
P_j = \sum_{M_i:\text{factor } j \text{ is active}} p(M_i|\boldsymbol{y}).
$$

A large value of $P_j$ indicates that factor $j$ is more likely to be active and a small value indicates that factor $j$ can be an inert factor. The probability $P_j$ was also used in Box and Meyer (1993), and they suggested that what is important is the pattern of these $P_j$'s not the values.

## An Example

In this section we analyze a data set from Box and Meyer (1993) to demonstrate our new approach and to compare with the result from original analysis. A 12-run Plackett-Burman design and the responses are shown in Table 1. It was originally constructed by extracting from a $2^5$ design which was from Box, Hunter and Hunter (1978). Traditionally, 11 effects are estimated first then the normal plot is used to identify significant effects. According to the result, only the main effect B is identified as significant. Box and Meyer (1993) observed from the alias relationships that 2fi effects BD and DE can be significant as well. They analyze the data again by using their Bayesian approach. The result shows that, in addition to factor B, both factor D and E are successfully screened out. For more details see Box and Meyer (1993).

Table 1: Responses for Example 1

| Run | A | B | C | D | E | Y |
|---|---|---|---|---|---|---|
| 1 | + | − | + | − | − | 56 |
| 2 | + | + | − | + | − | 93 |
| 3 | − | + | + | − | + | 67 |
| 4 | + | − | + | + | − | 60 |
| 5 | + | + | − | + | + | 77 |
| 6 | + | + | + | − | + | 65 |
| 7 | − | + | + | + | − | 95 |
| 8 | − | − | + | + | + | 49 |
| 9 | − | − | − | + | + | 44 |
| 10 | + | − | − | − | + | 63 |
| 11 | − | + | − | − | − | 63 |
| 12 | − | − | − | − | − | 61 |

We use our method to analyze the data with various choices of $\Delta$ and $\pi$. The integral in (4) is computed by using Monte Carlo methods with 1000 replicates. The result is shown in Figure 1. We can see that, first, the turning parameter representing the signal and noise ratio, and if a factor is active or not can be distinguish easily with a larger $\Delta$. Secondly, more factors will be identified as active with a larger $\pi$. In general, all the six plots are showing the same pattern and Factor B, D and E are identified as active, which is identical with Box and Meyer's result.

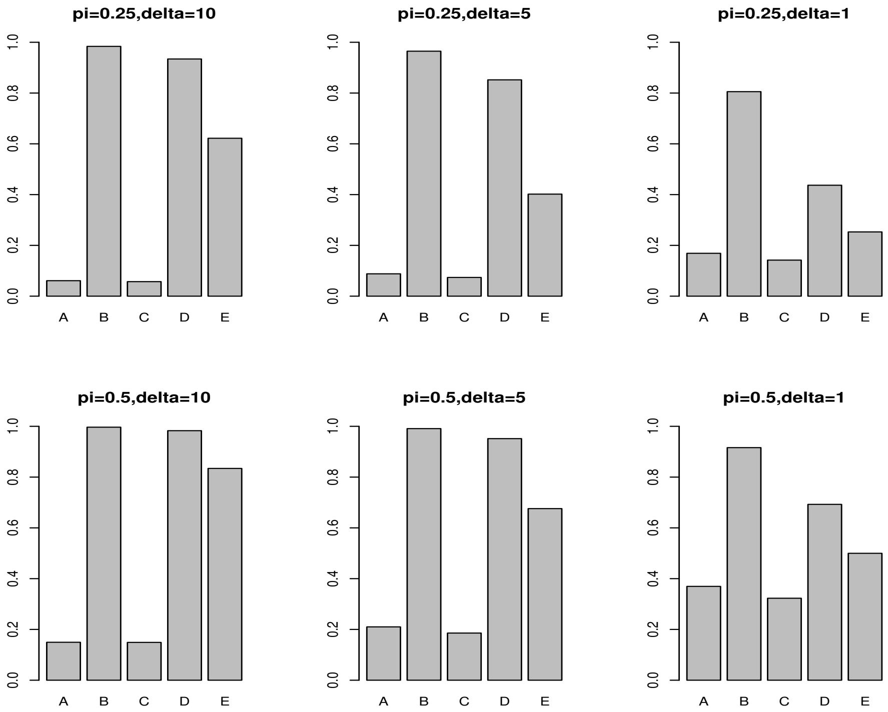


Figure 1: Posterior Probability of Active Factors

The most time-consuming part of using our method is to compute (4). If the integrals can be replaced by (5), then the computational time can be reduced dramatically. In Figure 2, we are showing the marginal posterior probabilities of active factors with different choices of $\rho$'s. We can see that it doesn't work well to identify active factos with larger $\rho$'s (0.9). However, the same pattern that posterior probabilities of factor B, D and E being active still stand out when $\rho$'s are small (0.1 and 0.5).

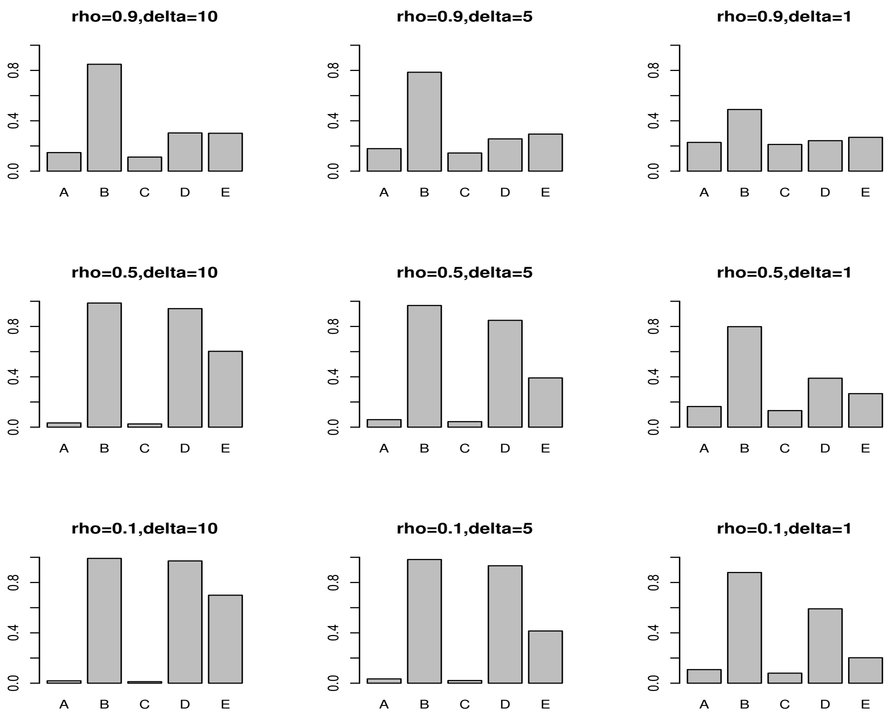


Figure 2: Sensitive Analysis of $\boldsymbol{\rho}$

## Conclusion

In this project, we propose a new factors screening method based on a GP under a Bayesian framework. This method considers only factors not effects and the total number of possible model considered can be reduced. Due to the limit of pages, we are only showing one example in this report from which we got the motivation. We have analyzed several data sets and our method works well especially when there are higher order interactions existing. We conclusion this report by the following suggestions in general. When the number of candidate factors are large, (5) is suggested with plugging in with smaller $\rho$'s, such as 0.1, to approximate the integral. For the choice of $\Delta$ and $\pi$, a larger $\Delta$ and smaller $\pi$ are suggested. However, a sensitive analysis to the choice of these parameter is still suggested.

## References


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

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for Experimenters*, John Weiley & Sons, New York, NY.

Kass, R.E. and Raftery, A.E. (1995). "Bayes Factors". *Journal of the American Statistical Association* 90, pp. 773-795.

Diggle, P.J. and Ribeior, Jr. P.J. (2007). *Model-based geostatistics*, Springer.

Santner, T.J., Williams, B.J., and Notz, W.I. (2003). *The design and analysis of computer experiments*, New York: Springer.

Wu, C.F.J. and Hamada, M. (2000). *Experiments: planning, analysis, and parameter design optimization*, New York: Wiley.

本人(俞一唐)於民國 98 年 2 月 4 日至 16 日按計畫赴美喬志亞理工學院工業工程系(School of Industrial and Systems Engineering in Georgia Institute of Technology)訪問 V.R. Joseph 博士。期間，針對貝氏模式選取在工業實驗上的應用進行實質上的探討,並討論將來合作的可能性。

Georgia Institute of Technology
H. Milton Stewart School of Industrial and Systems Engineering

November 24, 2008

Dr. I-Tang Yu
Department of Statistics
Tunghai University
Taichung, Taiwan

Dear Dr. Yu,

I would like to invite you to visit us during February 2-16, 2009. During this visit we will work on the Bayesian model selection problem and discuss the possibility of future collaboration. I hope you will accept my invitation.

Thank you,

Sincerely

V. Roshan Joseph
Coca-Cola Associate Professor

Atlanta, Georgia 30332-0205 U.S.A.
web site: http://www.isye.gatech.edu

Phone: 404•894•2300
Fax: 404•894•2301