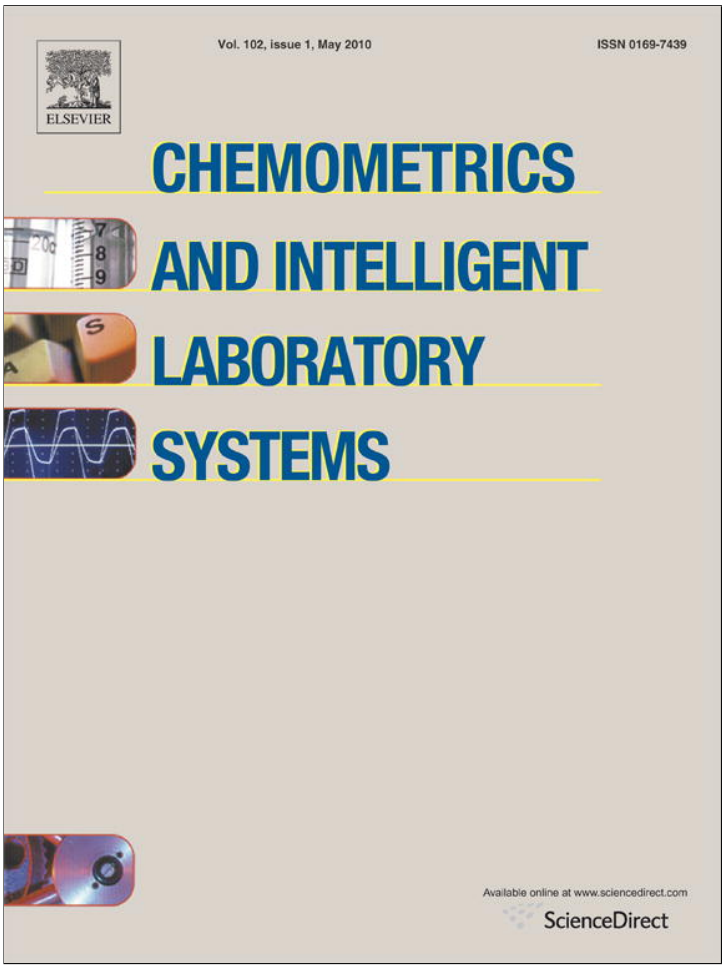

Chemometrics and Intelligent Laboratory Systems 102 (2010) 8–14

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

The effect of factor interactions in Plackett–Burman experimental designs Comparison of Bayesian-Gibbs analysis and genetic algorithms

Jorge F. Magallanes [a,*], Alejandro C. Olivieri [b]

[a] Comisión Nacional de Energía Atómica, Gerencia Química, Av. Gral. Paz 1499, B1650KNA San Martín, Pcia. de Buenos Aires, Argentina
[b] Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, and Instituto de Química Rosario (IQUIR), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Suipacha 531, Rosario S2002LRK, Argentina

ARTICLE INFO

Article history:
Received 13 December 2009
Received in revised form 9 February 2010
Accepted 22 February 2010
Available online 1 March 2010

Keywords:
Plackett–Burman designs
Factor associations
Genetic algorithms
Bayesian-Gibbs analysis
Experimental design and modeling

ABSTRACT

A genetic algorithm has been developed in order to estimate not only the main effects but also the association of terms when analyzing the influence of experimental factors through a Plackett–Burman design of experiments. The results for a series of simulated systems as well as experimental examples show excellent agreement with a Bayesian-Gibbs approach. The Plackett–Burman design is usually employed for screening, but its performance depends on the assumption that the interaction effects are negligible. Simulations allow one to analyze the effect of increasing interactions on the significance of main factors when Plackett–Burman designs are processed by neglecting factor associations.

1. Introduction

Experimental design helps chemometricians to select sample points for an experiment, all planned out in advance [1]. For this purpose, when one can freely select the ranges and levels for all variables, several designs have been developed whose main geometrical characteristic is the symmetric position of the samples in the multivariate space. Classical symmetrical designs are full factorial, fractional factorial, central composite, Box–Behnken, etc. [2–4]. Usually, the main purpose of an experimental design is to build a statistical model capable of estimating the behavior of the system under study (the response), starting from the factor values. These statistical models are mathematically described, in general, as a polynomial equation. A quadratic model to the analyzed response including k factors, when enough factor levels are available, is given by:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=i}^{k} \beta_{ij} x_i x_j + e, \quad (1)$$

where y is the system response, the values of β are coefficients to be estimated, x_i, x_j are the model factors, and e collects the model error. Eq. (1) includes an intercept (β_0), k main factor effects (β_i), $[k(k-1)/2]$ two-factor interaction terms (β_{ij}, $i \neq j$), and k quadratic terms (β_{ii}). Three-factor interaction terms are rarely needed, as models of order higher than two; in case of necessity it is more practical to shorten the variable ranges in order to describe the data using a simpler relationship of lower order.

An appropriate design will save time and effort by reducing the number of runs and decreasing the model errors to a minimum for a given number of experiments. Classical designs requiring only two levels for each variable, such as a two-level full factorial design, are only capable of describing linear relationships among the variables, as well as associations between different factors. For fully describing a second-order model [including the quadratic terms $\beta_{ii} x_i^2$ in Eq. (1)] it is necessary to have more than two levels for the relevant factors, and consequently the number of runs is increased.

A two-level full factorial design requires a total of 2^k runs to build a model, where k is the number of factors. One of the most important properties of the full factorial designs is its orthogonality, meaning that, from the point of view of regression, the variables are not correlated. As a consequence, the estimated β values do not influence each other. For more than three factors, the number of runs increases rapidly; due to this fact, fractional factorial designs have been devised, which save experiments by dividing the number of runs by 2 (or powers of 2). As a consequence, the possibility of independently estimating each model term is lost, and only estimations of confounded effects can be obtained [2,3]. In a general case, experimentalists may not suspect the existence of specific interactions, but may want to know which factors are significant and, at the same time, to verify whether some interactions could be significant or not, in order to validate the calculation of the main factors. This makes it difficult to set a specific fractional factorial design from the start.

* Corresponding author.
E-mail address: magallan@cnea.gov.ar (J.F. Magallanes).

0169-7439/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2010.02.007

An extremely economic design from the point of view of the number of experimental runs is the Plackett–Burman (PB) design, which is also a two-level orthogonal design [5]. A 12-experiment PB design, for example, allows one to study up to 11 factors. In comparison, a two-level full factorial design requires 32 runs for 5 factors, 64 for 6 factors, etc.

Considering two-level second-order models for evaluating only main factors and two-factor interactions, i.e., those given by the following expression:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j, \quad (2)$$

the number of PB runs is not enough to work out the complete system of equations for more than four factors (see below). As a consequence, PB designs can only estimate the main factors, and the terms taking into account the associations among them are confounded with the main effects or other associations, but not determined individually. The confounding pattern of the PB design is complex: every main factor is partially confounded with all possible two-factor interactions not involving the factor in question. Because of this fact, PB is described as a *screening* design. It is important to notice that the validity of a PB screening design to estimate the main factor effects depends on the assumption that the interaction effects are negligible, and thus such validity is often questionable [2]. The limits of this assumption have not been precisely established, and there exists the difficulty of not knowing the interactions beforehand. The subject is relevant since ignoring interactions may lead to the missing of important effects, to the incorrect consideration of irrelevant effects, or to the mistaking of effect signs, leading to wrong recommended factor levels.

Because PB designs are very appealing due to the apparent economy of experimental runs, it is stimulating to look for strategies that may allow practitioners to estimate the importance of interaction terms. In this case, PB designs could be used to build models such as those described by Eq. (2), besides its known utility as a screening design. Interactions can be brought out in PB designs, when the number of significant main effects is small, by: 1) regression analysis guided by the alias matrix [6], 2) the so-called frequentist approach and 3) Bayesian analysis using Gibbs sampling (BG) [1,7–14]. Very recently, these two latter approaches have been applied to some experimental systems by Wong et al., concluding that the BG technique is preferable for finding relevant associations when exploring factor influence using a PB design [15].

In the present report we describe an alternative strategy to BG analysis, based on a genetic algorithm (GA). Genetic algorithms are promising numerical optimization techniques which mimic natural selection processes [16–18], and can be used to obtain global minima for a complex multi-dimensional surface. In the context of experimental design they have not been fully explored, although they have been recently employed to aid in regression problems using supersaturated designs [19]. We have applied a suitably adapted GA, incorporating both the sparsity and heredity principles, to the analysis of simulated and experimental PB designs including two-factor interactions. The results to be described below are promising.

2. Strategies and methods

2.1. General considerations

As explained in the Introduction, the purpose of the present work is to find suitable models, based on the experiments performed according to a PB design, which are able to describe the most significant terms (main factors and two-factor interactions). We take as an example a 12-experiment, k-factor PB design. The complete design matrix includes the k columns corresponding to the main factors x_i ($i=1, 2,\ldots, k$), and the $[k(k-1)/2]$ columns corresponding to the interacting terms ($x_i x_j$) (for $i \neq j$). The total number of main and interacting terms is $[k(k+1)/2]$. For $k>4$, this latter number is larger than the number of runs required by the minimum PB design, and thus the design matrix is rank-deficient and cannot solve for all effects directly.

Classically, interacting terms are neglected, and the $12\times k$ design matrix **X**, with columns corresponding to the main factors only, is full-rank. It is employed to find the β coefficients through:

$$\beta = \mathbf{X}^{+}\mathbf{y} \quad (3)$$

where **y** is the vector of responses and the superscript '+' stands for the generalized inverse of a matrix. However, if interactions are present, the model may lead to erroneous results concerning the significance of the coefficients.

However, there are alternative models, which include some of the main factors and also some of their two-factor interactions, whose associated design matrices are full-rank, and therefore they allow for a direct least-squares solution of the corresponding model coefficients. Selection of these reduced models follows the assumptions that: 1) only a small number of factors will usually be significant (factor sparsity principle), and 2) significant interactions will occur between factors when at least one of them is in itself significant (heredity principle). The reduced models can be searched and guided by a significant improvement in model fit, i.e., a decrease in the root mean square error (RMSE), which is given by:

$$\mathrm{RMSE} = ||\mathbf{y}-\mathbf{X}\mathbf{X}^{+}\mathbf{y}|| / (\mathrm{DOF})^{1/2} \quad (4)$$

where **X** the design matrix including both main and interacting terms, and DOF the number of degrees of freedom, equal to twelve minus the number of β coefficients to be estimated. It is assumed that the reduced models fulfilling the above conditions will fit better than the classical, main factors only, PB design.

Methods for performing the above task are: 1) Bayesian-Gibbs analysis, 2) the frequentist approach and 3) the presently described genetic algorithm, which are described in the next sections.

2.2. The Bayesian method

The method based on Bayesian statistics and Gibbs sampling has been extensively described previously [2,7–11], and thus a brief description will be given here. A simple and reduced expression of Bayes theorem is the following [20]:

$$\mathrm{prob}(\mathit{hypothesis}|\mathit{data}, I) \;\alpha\, \mathrm{prob}(\mathit{data}|\mathit{hypothesis}, I) \times \mathrm{prob}(\mathit{hypothesis}|I) \quad (5)$$

This expression shows the power of the Bayes relationship, and relates the probability that the hypothesis be true given the data, to the probability that would have been observed with the measured data if the hypothesis was true. Although all the probabilities have been made conditional to I, which implies the relevant background information and assumptions one has about the system, it is often omitted. Some terms in the Bayes theorem have formal names, then, prob(*hypothesis*|*I*) is called the *prior* probability, and represent our state of knowledge about the trueness of the hypothesis before the current data have been analyzed. The prob(*data*|*hypothesis*,*I*) is the *likelihood function*, which modifies the experimental measurements. Both terms yield the *posterior* probability, prob(*hypothesis*|*data*,*I*), representing our state of knowledge about the trueness of the hypothesis in light of the data. A formal expression of the theorem, equating the previous expression and omitting I, is:

$$\pi(\theta|y) = \frac{f(y|\theta).\pi(\theta)}{\int f(y|\theta).\pi(\theta)d\theta} \quad (6)$$

where y denotes the data and θ denotes the hypothesis.

For our purpose, inference about θ can then be made using the posterior distribution. Consider a general linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{7}$$

where $\mathbf{X}$ is the model matrix, $\boldsymbol{\beta}$ is a vector containing the β parameters of Eq. (7) and $\boldsymbol{\varepsilon} \sim N(0,\sigma^2)$, is a vector of errors. For the variable selection strategy, θ is a vector $\boldsymbol{\theta}(\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2)$ where $\boldsymbol{\delta}$ specifies a model in which $\delta_i = 0$ for statistically non-significant terms ($\beta_i \rightarrow 0$) and $\delta_i = 1$ for the significant ones. Then, the priors for $\boldsymbol{\theta}$ in Eq. (6) are $\pi(\beta_i|\delta_i)$. The prior for σ^2, which is related to the probability of $\pi(\beta_i|\delta_i)$, is an inverted *gamma function* distribution.

As stated above, $\boldsymbol{\delta}$ specifies a model, and if each δ_i were independent, the prior for $\boldsymbol{\delta}$ would be:

$$\pi(\boldsymbol{\delta}) = \prod_{i=1}^{p+1} p_i^{\delta_i}(1-p_i)^{1-\delta_i} \tag{8}$$

where $p_i = \text{prob}(\delta_i = 1)$. In our problem, however, independent priors are not realistic and hence hierarchical priors are established according to the heredity principle. This principle establishes that the existence of an association (x_i x_j) is only taken into account when either x_i or x_j is present as significant factor.

The evaluation of the posterior in Eq. (6) is obtained using the Gibbs sampling method, which is a Markov Chain Monte Carlo technique [3 (p. 369)]. As a result of these calculations, a list is obtained relating each term of the general linear model [Eq. (7)] with the probability of its existence.

2.3. The frequentist approach and the genetic algorithms

In the frequentist approach, these three iterative steps are followed: 1) the significant main factors are found using a model which only includes linear terms in Eq. (2), 2) all possible two-factor interacting terms involving at least one of the significant effects found in the previous step are included in the model and the significant ones are identified by forward selection, and 3) the latter associations and all linear terms are included, and the significant main factors are identified by forward selection. The process is repeated until no statistically meaningful changes are found in the model fit.

A genetic algorithm represents a more elaborate and efficient version of the frequentist method. In a GA, a probabilistic approach inspired in natural selection mechanisms is applied, employing binary strings (chromosomes) containing genes which encode the experimental variables [21]. In the present case, an initial population is produced in the form of an ($X \times Y$) random binary matrix, where X is the total number of terms in Eq. (2) (main and two-factor associations), and Y is a predetermined number of chromosomes. For the operation of the heredity restriction, an association term is included if: 1) the corresponding gene encodes a '1' in the binary matrix, and 2) either of the corresponding main terms is included in the model as significant.

Once the Y different initial models are built, they are ranked according to a given objective function to be minimized. Half of the chromosomes having the best figures of merit are allowed to survive, mutate and recombine to produce offspring. When a single-point crossover is employed for chromosome recombination, a random point is first selected along a pair of parent chromosomes. The entire genetic information encoded in one of the parents up to the selected point is then transferred to the offspring, while the remaining genes are taken from the other parent (an alternative information transfer operates in creating a second offspring). After a number of generations in which the above scheme is repeated, the final best chromosome (i.e., the one leading to the minimum value for the objective function) is employed for model building.

In the present work, the GA was implemented starting with a population of 20 chromosomes, initialized with random binary digits with 20% probability for 1 and 80% for 0 values. The single crossover scheme with 50% probability was employed for recombination (the alternative multiple crossover procedure gave similar results), and a probability of 0.05 was applied to mutations after offspring were produced. The algorithm was stopped after 100 generations. The population size and number of generations were estimated by trial and error; for more complex systems (i.e., more active factors and experimental runs) correspondingly larger values than those herein employed might be required. Regarding the initialization scheme, other alternatives were tried (all values set at 1, and all values set at 0), in all cases with analogous results to those reported in the present paper. However, it should be noticed that these extreme alternatives led to initial populations where all individuals correspond to unacceptable solutions (rank-deficient when all values are set at 1 and empty when they are set to 0). Although the GA was able to find reasonable solutions, even under these extreme initialization conditions, it required more generations than the intermediate alternative.

The objective function to be optimized was the ratio of two RMSE values, one corresponding to the fit of the experimental responses to the current model and the second one to the fit of the classical PB analysis, i.e., Eq. (2) with only linear terms included (herein called RMSE_0). Before starting the GA, the responses where all converted to a scale with a minimum of 0 and a maximum of 1, and then mean-centered, which removes the need of the intercept β_0 in Eq. (2). All models which were not full-rank where automatically discarded by assigning an arbitrarily large RMSE value. The GA calculations were repeated ten times for each of the analyzed cases. A histogram was built registering the average value of coefficient terms over the ten GA cycles, and finally the terms having average coefficients larger than a certain tolerance (usually 0.05) were selected and included in a final model (in the case of the two-factor interactions, this was done provided they comply with the heredity principle). The least-squares coefficients solving for this model were normalized to unit length and reported with their associated standard deviations in order to judge their statistical significance. Specifically, the confidence interval for each coefficient was computed as [4]:

$$\text{CI}(\beta i) = t_{\nu,\alpha/2}\, s(\beta i) = t_{\nu,\alpha/2}\ \text{RMSE}\sqrt{\left[\left(X^T X\right)^{-1}\right]_{ii}} \tag{9}$$

where t is a Student coefficient for ν degrees of freedom (number of experiments minus number of estimated coefficients) and a $(1-\alpha)\times 100\%$ confidence level, and $s(\beta_i)$ is the standard error in the estimation of coefficient β_i. A term is not considered significant when $\beta_i \pm \text{CI}(\beta_i)$ includes the value of zero.

2.4. Comparison of models

The above mentioned methods for model evaluation rely on the improvement of the fit to Eq. (2). The predictive ability of different models can be judged from the distribution of prediction errors when they are used to predict a given response. In this sense, the RMSE is often used as a simple criterion for scoring predictive abilities of models. However, the question should be addressed whether selecting the model with minimum RMSE is indeed the best answer. Therefore, in order to check the significance of the assertion 'smaller RMSE values', a statistical F test and also the Van der Voet's randomization test based on t-test Monte Carlo simulations were applied [22]. In the present case the comparison involved the actual responses and those obtained by each of the different methods (Bayesian-Gibbs, frequentist or GA) and by the classical PB approach (i.e., neglecting interactions). For details on the Monte Carlo calculations, see Ref. [22]. As regards the F test, the ratio between both squared RMSE values was first computed, and the associated probability was found with the corresponding degrees of freedom (number of experimental runs minus number of adjustable parameters). In both cases, significance was established if the probability for the applied test was smaller than 0.05.

3. Results

3.1. Simulated systems

Two series of four models, including five main factors each, have been simulated. For the first of these models, the following expression was employed to simulate the responses:

$$y_1 = \beta_1 x_1 + \beta_3 x_3 + \beta_5 x_5 + \beta_{13} x_1 x_3 \quad (10)$$

The β coefficients were selected so that factors 1, 3 and 5 were significant, as well as an association between two significant main factors (1 and 3). Factors 2 and 4 were not considered significant, as all associations except 1–3. In the four y_1 models, the values of β_1, β_3 and β_5 were kept constant at 1, −1 and 1 respectively, while β_{13} was given the values 1, 2, 3 and 4. The vector of coefficients β was then normalized (see systems 1–4 in Table 1), and the models were analyzed at two-factor levels (−1 and 1), according to a PB design of 12 runs. Once the twelve responses were calculated, they were scaled (in the range of 0–1) and random Gaussian noise $N(\mu,\sigma^2)$, $\mu=0$, $\sigma=0.05$ was added. Then the models were evaluated in order to assess the terms recovered as statistically significant. Three different methods were used to work out the models: direct estimation of effects through critical errors from PB chart, the GA method and the Bayesian method.

Table 1
Results for the eight simulated systems studied with standard Plackett–Burman (PB), Bayesian-Gibbs (BG) and genetic algorithm (GA) analyses.

Term	β	PB	GA	BG	β	PB	GA	BG
	System 1				System 2			
x_1	0.50	Yes	0.48	1.000	0.38	Yes	0.47	0.873
x_2	0.00	No	–	<0.001	0.00	No	–	0.009
x_3	−0.50	Yes	−0.47	1.000	−0.38	Yes	−0.36	0.919
x_4	0.00	No	–	<0.001	0.00	No	–	<0.001
x_5	0.50	Yes	0.49	1.000	0.38	**No**	0.39	0.871
$x_1 x_3$	0.50	–	0.55	1.000	0.76	–	0.71	0.919
	System 3				System 4			
x_1	0.29	**No**	0.28	0.734	0.23	**No**	0.21	0.462
x_2	0.00	No	–	0.015	0.00	No	–	<0.001
x_3	−0.29	**No**	−0.28	0.871	−0.23	**No**	−0.19	0.677
x_4	0.00	No	–	0.023	0.00	No	–	0.003
x_5	0.29	**No**	0.28	0.364	0.23	**No**	0.21	0.062
$x_1 x_3$	0.87	–	0.87	1.000	0.92	–	0.93	1.000
	System 5				System 6			
x_1	0.50	Yes	0.50	1.000	0.38	Yes	0.38	0.999
x_2	0.00	No	–	<0.001	0.00	No	–	0.001
x_3	0.50	Yes	0.49	1.000	0.38	**No**	0.40	0.999
x_4	0.00	No	–	0.003	0.00	No	–	0.008
x_5	−0.50	Yes	−0.52	0.998	−0.38	**No**	−0.43	0.992
$x_1 x_2$	0.50	–	0.48	0.998	0.76	–	0.71	1.000
	System 7				System 8			
x_1	0.29	**No**	0.32	1.000	0.23	**No**	0.27	1.000
x_2	0.00	No	–	<0.001	0.00	No	–	<0.001
x_3	0.29	**No**	0.29	0.651	0.23	**No**	0.23	0.122
x_4	0.00	No	–	0.004	0.00	No	–	<0.001
x_5	−0.29	**No**	−0.30	0.670	−0.23	**No**	−0.23	0.150
$x_1 x_2$	0.87	–	0.85	1.000	0.92	–	0.91	1.000

The β values indicate the original coefficients employed in Eq. (10) to simulate the responses, PB the results of standard Plackett–Burman analysis (incorrect results in boldface), GA the significant coefficients provided by the genetic algorithm, and BG the Bayesian-Gibbs probabilities associated to each term. For BG calculations, the following parameters were used: iterations, 5000, inverse Wishart parameters, 2 and 0.01, prior for main effects, 0.1, probability for the interaction *ij* if both *i* and *j* are present, 0.5, probability for the presence of *ij* if either *i* or *j* is present, 0.5, probability for the presence of *ij* if neither *i* nor *j* are present, 0.00001.

Table 1 shows the results using classical PB analysis: the PB column indicates with 'Yes' the main factor effects that are higher than the critical error, which was calculated as:

$$C_{err} = t_{d,\alpha/2}||\mathbf{e}_d|| / d \quad (11)$$

where $t_{d,\alpha/2}$ is as in Eq. (9), $\mathbf{e}_d$ is the vector of effects of the dummy factors and d is the number of dummy factors [2, pp. 697–698].

The results of Table 1 show that for systems 1–4, increasing the relative weight of the interaction β_{13} leads to increasingly wrong performances of the PB evaluation. Moreover, the correlation coefficient r^2 decreases from ca. 1 when the interaction is absent to ca. 0.40 for the maximum β_{13} studied, indicating how the fit is degraded when strong two-factor interactions are present. Fig. 1A shows the progression of the five standard PB coefficients and the critical error for increasing values of β_{13} in the studied range (normalized values are shown in Fig. 1A). The shaded area in Fig. 1A is bounded by the critical error [Eq. (11)], with significant terms having β values outside this region. Fig. 1A shows the standard PB evaluation of the main factors starts to fail when the association term β_{13} is ca. 0.55, i.e., a squared weight of about 30% (computed with reference to the sum of squared β coefficients, which is equal to unity). Fig. 1A also shows a progressive decrease of the estimated value for β_5, indicating how this term looses significance when PB analysis is applied without taking into account the $(x_1 x_3)$

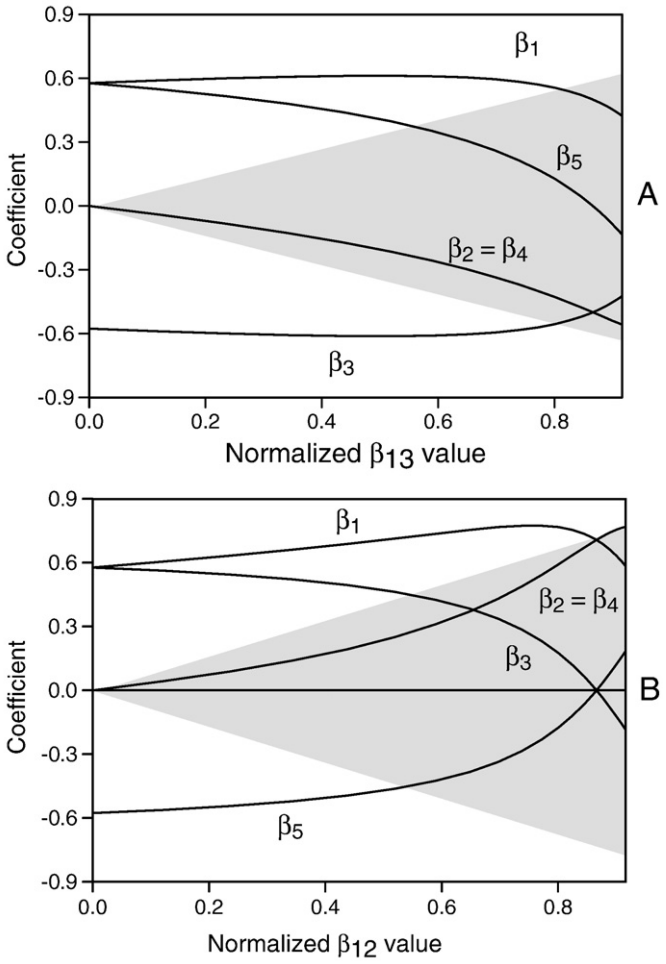

Fig. 1. A) Changes in the values of the five coefficients (corresponding to the five main factors) obtained by standard PB analysis of the simulated system 3, as a function of increasing interaction between factors 1 and 3 [normalized β_{13} values in Eq. (10)]. B) Same as A) for system 7, as a function of increasing interaction between factors 1 and 2 [normalized β_{12} values in Eq. (12)]. The shaded area is bounded by the critical error values [see Eq. (11)]. Coefficients outside the shaded area are considered as significant.

interaction. Moreover, beyond the limit $\beta_{13} = 0.85$ the sign of the coefficient is reversed. This result can be explained by the relationship among coefficients in this particular system. If β_{50} is the coefficient corresponding to the linear term x_5 when standard PB analysis is applied, and β_5 and β_{13} are the estimated coefficients for a model described by Eq. (10), then it can be shown that:

$$\beta_{50} = \beta_5 - \beta_{13} / 3 \tag{12}$$

which can be traced to the fact that (x_1x_3) is partially confounded with all linear terms except 1 and 3, i.e., with x_2, x_4 and x_5 (see Ref. [2], p. 351). Since only x_5 is significant, the correlation between the interaction (x_1x_3) and the main term x_5 is:

$$(\mathbf{x}_1 \cdot \mathbf{x}_3)^T \mathbf{x}_5 = -1 / 3 \tag{13}$$

where $\cdot$ is the Hadamard or element-wise product operator. This explains the $(-1/3)$ factor in Eq. (12) in front of β_{13}, which shows that under strong β_{13} interactions, β_{50} tends to decrease, up to a point where it can be (incorrectly) considered as non-significant (Fig. 1A).

It is also worth noting that the importance of both β_2 and β_4 increases with increasing interaction between factors 1 and 3, with the concomitant risk of misinterpreting the spurious factors 2 and 4 as significant (Fig. 1A).

It is interesting to note that the frequentist approach mentioned in Section 2.3. cannot be applied to some of the systems in Table 1, e.g., those in which none of the standard PB coefficients are significant (i.e., systems 3, 4, 7 and 8). However, in all cases cited in Table 1, application of the GA method allows one to significantly improve the fit, by combining main factors and associations fulfilling both the factor sparsity and the heredity principles. Fig. 2A shows how the ratio (RMSE/RMSE$_0$) decreases for a typical GA run corresponding to system 3 in Table 1, while Fig. 2B displays the final results in the form of a histogram, where the significance of the average coefficients can be graphically gathered. For the analysis of this particular system 3, the value of r^2 increases from 0.372 (standard PB analysis) to 0.982 (final GA model), while the RMSE decreases from 0.33 (standard PB analysis) to 0.051 (final GA model), the latter being nicely compatible with the degree of response noise introduced in the system (0.05 units). Furthermore, GA allows one to find a full-rank model whose expression is identical to the response generating expression (10), i.e., including only factors 1, 3 and 5, and the association 1–3.

Table 1 collects the significant model terms according to GA calculations, in the form of normalized values for the coefficients for all systems. As can be seen, correct solutions were found in the simulated cases using GA analysis, including the finding of the significant interacting term (x_1x_3), apart from the significant main factors (x_1, x_3 and x_5). Agreement was found in the absolute values of the coefficients for the significant terms, and also in their signs (see Table 1). The comparison of the RMSE values with the standard PB RMSE$_0$ value using both the F test and the randomization test discussed above led to a significant improvement in all cases ($p<0.05$). It is important to remark that a model such as that described by Eq. (2), which hypothetically has 15 different coefficients, has been analyzed with only 12 runs.

The results from BG analysis are also quoted in Table 1, in the form of posterior probabilities associated with the different terms. The results agree with those provided by GA. However, it may be noticed that BG probabilities associated with the important factor x_5 appear to be rather small for the strong interacting systems 3 and 4 (Table 1).

The second series of four models was generated from the following expression:

$$y_2 = \beta_1 x_1 + \beta_3 x_3 + \beta_5 x_5 + \beta_{12} x_1 x_2 \tag{14}$$

In these systems, the values of β_1, β_3 and β_5 were kept constant at 1, 1 and −1 respectively, while β_{12} was given the values 1, 2, 3 and 4. Then a similar approach to that discussed above for y_1 models was applied. Notice that in this model the interaction takes place between a significant (x_1) and a non-significant (x_2) main factor. Applying the heredity principle should allow the different methods to find this interaction.

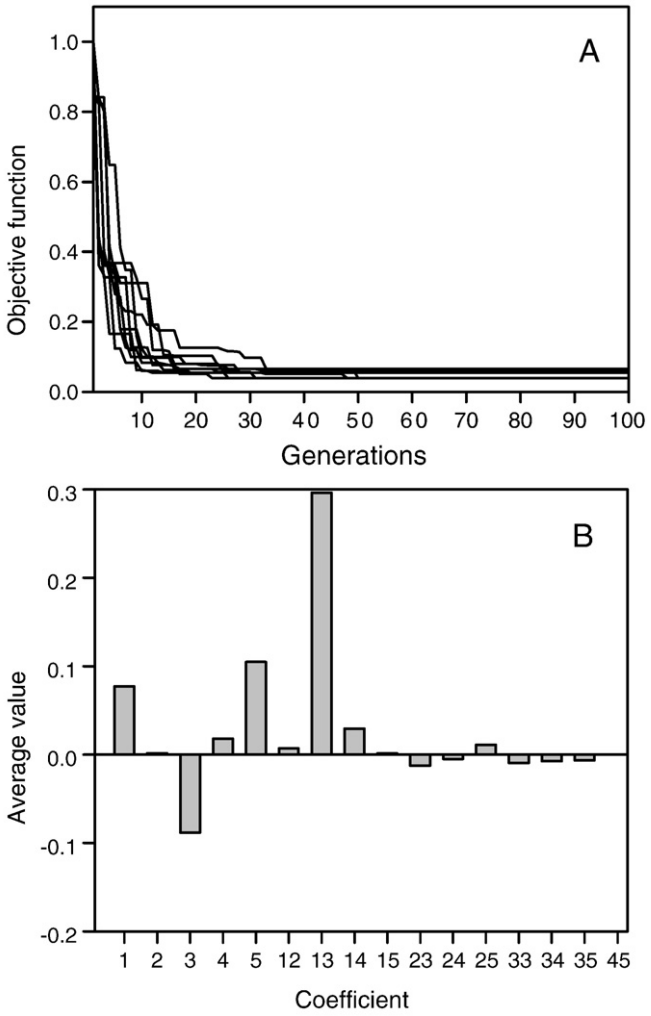

Fig. 2. A) Evolution of the objective function with the generations, during the ten runs of the genetic algorithm which corresponds to system no. 3 in Table 1. The objective function is the ratio of RMSE for the current model and the RMSE$_0$ for the standard PB analysis. B) Histogram registered after calculations with the genetic algorithm corresponding to the same system. The bars represent the average coefficients found after ten GA cycles.

Standard PB evaluation leads to the results quoted in Table 1. As before, increasing the interaction leads to increasingly incorrect results regarding the significance of the main factors. Fig. 1B shows the changes in the standard PB coefficients and critical error for increasing the values of β_{12}. The standard PB evaluation starts to fail when the association term β_{12} has a squared weight of about 30%. Fig. 1B also shows a progressive loss of significance for the important factors β_3 and β_5, and, beyond the limit $\beta_{12} = 0.85$, a reversal in the signs of these two coefficients. Interestingly, the irrelevant factors β_2 and β_4 appear to gain importance as β_{12} increases, which may lead them to be confused with important effects.

GA analysis leads to correctly identifying the main factors and the relevant interaction, as can be seen in Table 1. Again, the signs and values of the coefficients agree well with those employed to generate the simulated responses. The least-squares fit did also significantly improve in comparison with standard PB analysis, leading to all quoted coefficients in Table 1 to be significant. Finally, the comparison of the RMSE values with the standard PB RMSE$_0$ value using both statistical tests discussed above led to significant improvement in all cases.

In the case of BG analysis, as can be seen in Table 1, general agreement is found with the use of the GA, except perhaps in systems with strong 1–2 interactions, where the contribution from the significant main terms 3 and 5 is found to be comparatively smaller than x_1 and (x_1x_2).

3.2. Real systems

3.2.1. Cast fatigue experiment

An experimental example concerning cast fatigue has been discussed by Wu [2], and originally by Hunter et al. [23]. In this experiment, seven factors (named *A* to *G*) have been studied as main factors by Hunter through a PB design of 12 runs. Standard PB analysis gives *F* as the only significant main factor. The simple model implied by this result, i.e.:

$$Z = 5.73 + 0.46F \quad (15)$$

where *Z* is the estimated response, gives an RMSE of 0.55 units (ca. 10% with respect to the mean response) and $r^2 = 0.67$. Wu did also consider the interactions using the same data set, concluding that the significant terms were the 6th (*F*) both as main factor and as an interaction with the 7th factor (*G*), i.e., *FG* [2]. We analyzed again these results using both BG and GA methods. As we remarked previously, in this case the number of hypothetical terms of a model obeying Eq. (2) is larger than the number of PB runs.

The application of BG analysis confirms the result obtained by Wu, with a probability for the *F* factor as 0.973, and for the interaction *FG* as 0.951. The probability for the main term *D*, which was carefully considered by Wu, was 0.065 and all the hypothetical two-factor interaction terms have probabilities of less than 0.05 (readers interested in further details may refer to Ref. [2]).

The outcome of the GA method is in agreement with these previous results. The fit to models which include interactions among factors considerably improves in comparison with the standard PB analysis, with the largest coefficients corresponding to the terms *F* (0.72) and *FG* (−0.66) (normalized β values are given in parenthesis), both of which are significant. Furthermore, a small contribution from *D* was also found (the normalized β is −0.19), which does also appear to be significant.

A model including only *F* and *FG* terms leads to the following expression:

$$Z = 5.73 + 0.46F - 0.46FG \quad (16)$$

yielding RMSE = 0.26 units (4.5%) and $r^2 = 0.945$. This implies an important improvement in fit with regard to traditional PB analysis.

On the other hand, including the smaller term *D* leads to the following model:

$$Z = 5.73 - 0.17D + 0.46F - 0.46FG \quad (17)$$

which provides RMSE = 0.20 units (3.5%) and $r^2 = 0.979$.

Application of Van der Voet's randomization test leads to the conclusion that the model shown in Eq. (16) is indeed significantly better than standard PB analysis ($p = 5\times10^{-4}$). However, comparison of the RMSE furnished by the model expressions (16) and (17) gives $p = 0.19$, indicating that the latter is not significantly better than the former. Hence, the inclusion of *D* does not imply a statistically significant better fit.

Using *F* test considerations, similar conclusions are reached: while the RMSE value for Eq. (16) is significantly smaller than for Eq. (15) ($p = 2\times10^{-3}$), the comparison of RMSE values for Eqs. (16) and (17) implies that they are statistically undistinguishable ($p = 0.17$).

3.2.2. Adsorption of arsenic by biogenic hydroxyapatite

A study about the arsenic adsorption by biogenic hydroxyapatite in water has been carried out by Czerniczyniec et al., with the intention of developing a technique for reducing the concentration levels of As in drinking water [24]. The influence of five factors on the adsorption performance was analyzed using a PB design of 12 runs. The factors were: pH (*H*), water hardness (*I*), alkalinity (*J*), calcination temperature of bones (the source of biogenic hydroxyapatite, *K*) and salinity of water (*L*). Table 2 shows the PB experimental chart. When applying standard PB analysis, salinity (*L*) appears to be the only significant factor, leading to RMSE = 90 ppm (11%) and $r^2 = 0.672$, i.e., a rather poor fit.

Table 2
PB design of the adsorption of arsenic by biogenic hydroxyapatite in water.

Factor	Unit	Level +	Level−	Run	*H*	*I*	*J*	*K*	*L*	Response
pH		8.6	6	1	+	−	+	−	−	892
Hardness	ppm	180	30	2	+	+	−	+	−	995
Alkalinity	ppm	250	90	3	−	+	+	−	+	722
Temp. calc.	°C	900	500	4	+	−	+	+	−	952
Salinity	M	0.5	0.05	5	+	+	−	+	+	670
				6	+	+	+	−	+	736
				7	−	+	+	+	−	805
				8	−	−	+	+	+	744
				9	−	−	−	+	+	712
				10	+	−	−	−	+	736
				11	−	+	−	−	−	650
				12	−	−	−	−	−	926

H = pH, *I* = hardness, *J* = alkalinity, *K* = temperature of calcination, *L* = salinity. Response = As concentration in ppm.

If the frequentist approach is applied to this experimental system, one starts by finding the only significant effect by standard PB analysis, which is *L* (step 1). Then the fit to models having *L* and all possible two-factor associations including *L* are considered (step 2). However, this leads to four models, all having a similar poor fit, meaning that in this case the frequentist approach is not able to find any relevant interaction.

The application of the GA approach to this system provides interesting results concerning factor associations. Table 3 shows that the most significant coefficient found by GA is *L*. The interactions *HI* and *KL* are also considered important by the GA. Additional terms and interactions such as *I*, *HL*, *IK* and *IL* do also appear to play a role. Including the significant GA terms leads to the model:

$$W = 795.0 - 39.2I - 82.3L + 49.2HI - 37.2HL + 27.3IK + 21.3IL - -58.8KL \quad (18)$$

where *W* is the estimated response. This gives an improved fit, because RMSE decreases to 15 ppm (2%) and $r^2 = 0.997$ are obtained. The randomization test indicates that the fit improvement by using Eq. (18) is significant over standard PB analysis ($p = 2.5\times10^{-3}$). An analogous conclusion is reached using the *F* test ($p = 2\times10^{-3}$). This implies that factors *I*, *K* and *H* should be considered for subsequent research phases (i.e., optimization), a result which does not follow from standard PB analysis. The weights of the significant association terms provided by GA results may help to explain the unsuccessful

Table 3
Results obtained using Bayesian-Gibbs (BG) and genetic algorithm (GA) analysis for the five factors influencing the arsenic adsorption by biogenic hydroxyapatite in water.

Main factors	Term	BG probability	GA coefficient
pH	*H*	0.302	–
Hardness	*I*	0.419	−0.34
Alkalinity	*J*	0.041	–
Calcination temperature	*K*	0.095	–
Salinity	*L*	0.826	−0.62
	HI	0.404	0.35
	HJ	0.010	–
	HK	0.014	–
	HL	0.265	−0.30
	IJ	0.048	–
	IK	0.061	0.25
	IL	0.053	0.17
	JK	0.008	–
	JL	0.135	–
	KL	0.150	−0.45

BG parameters as in Table 1.

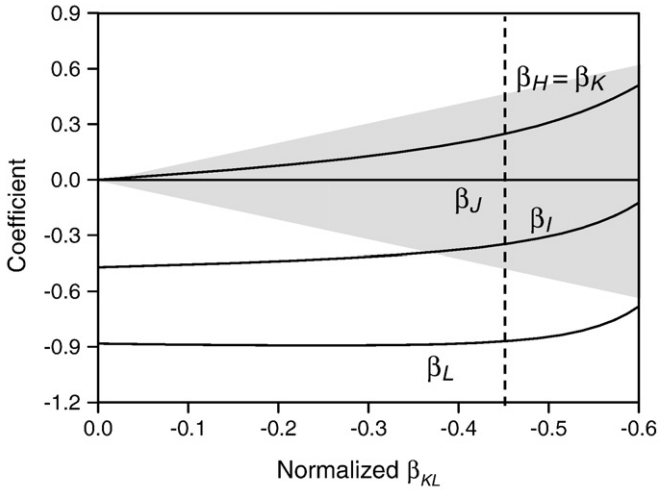

Fig. 3. Changes in the values of the five coefficients obtained by standard PB analysis of a system similar to the experimental one concerning arsenic adsorption, as a function of increasing interaction between factors K and L (normalized value of β_{KL}). The shaded area is bounded by the critical error values [see Eq. (12)]. Coefficients outside the shaded area are considered as significant. The dashed line marks the value of β_{KL} obtained from GA analysis.

performance of PB to this case. Fig. 3 shows the changes in standard PB coefficients which are expected for a hypothetical system having $\beta_I = \beta_L/2$ (both negative), and increasing values of the principal associations $\beta_{KL} = -\beta_{HI}$ (β_{HI} positive), i.e, similar to the presently studied one. As can be seen, the standard PB value for term I looses significance when the normalized interaction coefficient is larger than 0.38, which is attained in the experimental system (Table 3).

Table 3 shows the corresponding BG results for this design, which reveal that salinity (L) is the most important main factor, but pH (H) and hardness (I) are also significant. Furthermore, two relevant interactions appear for the pair of factors pH-hardness (HI) and pH-salinity (HL), as well as two additional, minor interactions (JL and KL). A model including all these terms leads to:

$$W = 795.0 + 18.9H - 42.3I - 91.3L + 48.9HI - 27.1HL - 4.6JL - 53.4KL \quad (19)$$

which furnishes RMSE = 50 (6.3 %), $r^2 = 0.934$, indicating an apparent worse fit in comparison with the GA model, a fact which is confirmed by application of the randomization test for the comparison of RMSE values ($p = 2.5 \times 10^{-3}$) and also by the F test ($p = 2 \times 10^{-3}$).

This is a good example where BG and GA can find not only the main factors hidden by the associations, but the associations themselves. Overall, however, the results tend to favor the GA calculations for this system. The present results are confirmed by a previously published report where a Doehlert design with a larger number of factor levels was processed with an artificial neural network [24]. This latter work showed that salinity (L) is the major variable affecting the efficiency of the arsenic immobilization process, with the following significant interactions: pH-salinity (HL), pH-hardness (HI), hardness-salinity (IL) and hardness-calcination temperature (IK). These more elaborate results, requiring a larger number of experiments, are in good agreement with those furnished by GA analysis of the simpler, 12-experiment PB design.

4. Conclusions

A study about the possibility of using the Plackett–Burman experimental design to build models that include associated terms has been carried out. In this regard, Bayesian-Gibbs analysis provides an efficient tool for estimating the significant terms, while an alternative genetic algorithm has the ability to provide the values of the model coefficients directly. The agreement in the estimation of terms between these two latter techniques is satisfactory. Regarding the use of the Plackett–Burman design for screening, it seems very risky to directly apply it for estimating main factors without the knowledge of whether the associated factors are negligible or not. This inconvenience can be solved by applying any of the presently discussed methods (Bayesian-Gibbs or genetic algorithms) to the same data set. From this point of view, the Plackett–Burman design could not only be considered as a screening design, but as a design which allows one to build models with a great economy of runs, provided it is complemented with the appropriate approaches for uncovering factor interactions.

Acknowledgments

The University of Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, project no. PIP 1950) and the ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, project no. PICT 25825) are gratefully acknowledged for the financial support.

References

[1] R. Cela, Screening Strategies, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Vol. 1, Elsevier, Amsterdam, 2009, p. 251.
[2] J.C.F. Wu, M. Hamada, Experiments, Planning, Analysis, and Parameter Designs Optimization, Wiley-Interscience Publications, New York, 2000 (Chapter 8).
[3] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997 (Chapters 21–24).
[4] G.E.P. Box, W.G. Hunter, J.S. Hunter, Statistics for Experimenters, John Willey & Sons, New York, 1978.
[5] R.L. Plackett, J.P. Burman, Biometrika 33 (1946) 305–325.
[6] J. Lawson, Comput. Stat. Data Anal. 39 (2002) 227–241.
[7] M. Hamada, C.F.J. Wu, J. Qual. Technol. 24 (1992) 130–137.
[8] H. Chipman, M. Hamada, C.F.J. Wu, Technometrics 39 (1997) 372–381.
[9] P.J. Brown, M.J.R. Vannucci, J.R. Stat. Soc. B 60 (1998) 627–641.
[10] M. Vannucci, MATLAB Code for Bayesian Variable Selection, The ISBA Bulletin, Vol. 7, 2000, (September, n. 3).
[11] www.cs.brown.edu/research/ai/dynamics/tutorial/Documents/GibbsSampling.html.
[12] D.K.J. Lin, N.R. Draper, Technometrics 34 (1992) 423–428.
[13] C.S. Cheng, Ann. Stat. 23 (1995) 1223–1233.
[14] C.S. Cheng, Biometrika 85 (1998) 491–495.
[15] F.K.H. Phoa, W.K. Wong, H. Xu, J. Chemometr. 23 (2009) 545–553.
[16] L. Davis, Genetic Algorithm and Simulated Annealing, Pitman, London, 1987.
[17] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1988.
[18] R.J. Leardi, J. Chemometr. 14 (2000) 643–655.
[19] R. Cela, E. Martínez, A.M. Carro, Chemometr. Intell. Lab. Syst. 57 (2001) 75–92.
[20] D.S. Sivia, Data Analysis, A Bayesian Tutorial, Oxford University Press, Oxford, UK, 1998.
[21] R.E. Shaffer, G.W. Small, Anal. Chem. 69 (1997) 236A–242A.
[22] H. van der Voet, Chemometr. Intell. Lab. Syst. 25 (1994) 313–323.
[23] G.B. Hunter, F.S. Hodi, T.W. Eager, Metall. Trans. 13A (1982) 1589–1594.
[24] M. Czerniczyniec, S. Farias, J.F. Magallanes, D. Cicerone, Water Air Soil Pollut. 180 (2007) 75–82.