# PARETO SIMULATED ANNEALING FOR THE DESIGN OF EXPERIMENTS: ILLUSTRATED BY A GENE EXPRESSION STUDY

Penny SANCHEZ [*], Gary GLONEK [†], Andrew METCALFE [†]

**Abstract.** Experimental design is concerned with the problem of allocating resources within an experiment to ensure that objectives of the experiment are achieved at the minimum cost. This paper focuses on the generation of optimal or near-optimal designs for large and complex experiments where it is infeasible to carry out an exhaustive search of the design space. Optimal designs for gene expression studies, aimed at investigating the behaviour of genes, are considered, where the optimality criterion employed is Pareto optimality. We develop an adaptation of the metaheuristic method of Pareto simulated annealing to generate an approximation to the set of Pareto optimal designs for large and complex experiments. We develop algorithms that utilise response surface methodology to search systematically for the optimal values of parameters associated with Pareto simulated annealing and performance is evaluated using quality measures.

**Keywords:** Combinatorial optimisation, Metaheuristics, Pareto simulated annealing, Experimental design, Gene expression study

## 1 Introduction

Experimental design is concerned with the problem of allocating resources within an experiment to ensure that objectives of the experiment are achieved at the minimum cost. The objectives can be modelled such that effects of interest are identified and any constraints can be taken into account. Rigorous experimental design is essential to make the most effective use of available resources and should be considered prior to carrying out the experiment.

[*]University of South Australia, School of Mathematics and Statistics, Mawson Lakes Campus, Mawson Lakes, South Australia, 5095, Australia

[†]University of Adelaide, School of Mathematical Sciences, Adelaide, South Australia, 5005, Australia

Suppose some system response, denoted by say $m$, is thought to depend on the values taken by predictor variables, denoted by $x_1, x_2, ..., x_p$. The relationship is approximated by a mathematical model which includes unknown coefficients. The aim is to estimate the values of these coefficients from runs of the process which result in a value of the response for corresponding choices of the value of the predictors. The experiment consists of a set of runs and the design is the choice of values of the predictor variables for these runs. The choice of the values of the predictor variables affects the precision of the estimates and objectives are to minimise the variance of estimators and the variance of specified combinations of estimators. Under rather general conditions these objectives are independent of the values taken by the response and the choice of design can be made on a quantitative basis before the experiment is performed.

In many practical problems, the predictor variables may be restricted to a few, or even just two, categories. In this case the problem of choosing a design is a combinatorial problem. We consider the case of a gene expression study. Such studies are aimed at investigating the behaviour of genes and provide the potential to make significant advances in areas of medical research, including the prevention, treatment and cure of genetic conditions (see [11], [6] and [8] for further background about gene expression studies).

Consider a gene expression study aimed at investigating the genetic basis of leukaemia (see [11] for further background). In this gene expression study, there are two factors, each with two categories: cell line, which can be normal (wild type) or mutant; and time which is no delay or 24 hours delay. It follows that there are 4 combinations: wild type with no delay, wild type with delay, mutant with no delay and mutant with delay. An experimental run involves comparing two of the four possible combinations on a microarray slide. Since there are 6 ways of choosing 2 from 4, there are 6 slide types, which constitutes the set of configurations for the experiment. The possible comparisons are shown in Figure 1 The two combinations on a microarray slide are dyed red and green, more specifically referred to as two-colour microarrays, and the response is the relative log expression which ranges from green through yellow to red. The design is a specification of the number of slides of each type, or configuration, given a fixed number, $n$, of available slides. A design can be represented by the sextuple $\mathbf{d} = (d_1, d_2, ..., d_6)$, such that

- $d_1, d_2, ..., d_6 \geq 0$,

- $d_1 + d_2 + ... + d_6 = n$.

At the outset of the gene expression study, the effects of scientific interest are identified and prioritised, the number of available microarray slides is fixed and an appropriate mathematical model is adopted. (See [14] for a general introduction to design issues for such studies.) The appropriate mathematical model relates the combinations or pairs of samples chosen for each available microarray slide, or design, to the efficiency of effects that are of scientific interest in the experiment.

Our optimality criterion is Pareto optimality, as in [11] and [5]. Pareto optimality is an appropriate optimality criterion because analysis is carried out separately for the effects of particular scientific interest for gene expression studies that use two-colour
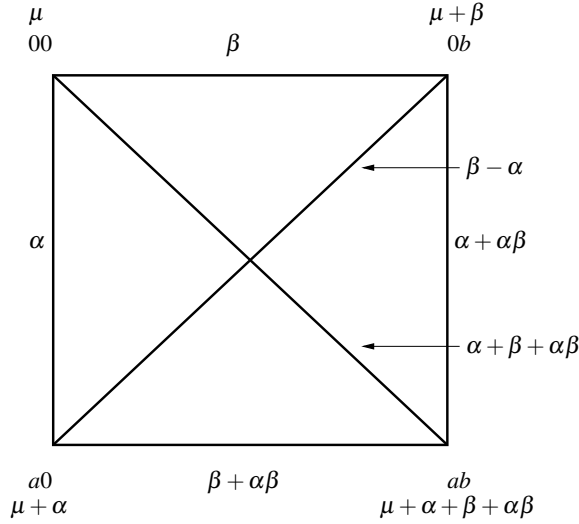
**Figure 1**: Parameterization and hybridizations for the $2 \times 2$ experiment.

microarrays. For further details regarding gene expression analysis, an introduction is in [13]. For further details on the gain in efficiency using Pareto optimality over other classical optimality criteria and associated designs in the two-colour microarray context, see [5].

A design is defined to be Pareto optimal if there is no other design that leads to equal or greater precision for each effect of scientific interest and strictly greater precision for at least one. The combinatorial nature of the problem lies in the choice of optimal designs, that is, the combinations of experimental material that are to be applied to each available microarray slide such that the effects of particular scientific interest are optimised. Such an approach explores the trade-off in efficiency for the effects of scientific interest for various designs.

In what follows, the background underpinning the formation of suitable objectives to which Pareto optimality can be applied is given. Firstly, in general a linear model for the set of configurations can be represented by

$$\boldsymbol{\theta} = A\boldsymbol{\beta} \tag{1}$$

where $\boldsymbol{\theta}$ is the vector of expected log ratios, $A$ is the configuration matrix and $\boldsymbol{\beta}$ is the parameter vector.

For the present gene expression study, $\boldsymbol{\beta}$ is $(\alpha, \beta, \alpha\beta)$, representing the change in expression between the wildtype and mutant with no delay, the change in expression for the wildtype over time and the interaction between cell line and time respectively

and

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

The observed log-intensities in the experiment, $\boldsymbol{M}$, are related to the parameters through the design matrix $X$, the rows of X being the rows of A corresponding to the choice of configurations used in the experiment. Thus

$$E(\boldsymbol{M}) = \boldsymbol{\eta} = X\boldsymbol{\beta} \tag{2}$$

and

$$\mathrm{var}(\boldsymbol{M}) = \sigma^2 I \tag{3}$$

where $\boldsymbol{M}$ is the vector of log intensity-ratios from all slides in the experiment, $\boldsymbol{\eta}$ is the vector of expected values, $X$ is the design matrix and $\boldsymbol{\beta}$ is the parameter vector.

The least squares estimators of $\boldsymbol{\beta}$ are given by

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{M}$$

and the variance-covariance matrix is given by

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (X^T X)^{-1},$$

see, for example, [12].

However, in this, and similar, gene expression studies some linear combinations of parameters, known as contrasts, are also of interest. Using the linear model (1), a set of contrasts, say $\boldsymbol{\gamma}$, can be represented by a set of linear functions of $\boldsymbol{\theta}$. Let the $i$-th contrast be given by $\gamma_i = \boldsymbol{b}_i^T \boldsymbol{\theta}$ where $\boldsymbol{b}_i$ is a suitable vector of coefficients, and write $\boldsymbol{\gamma} = B\boldsymbol{\beta}$ is the complete set of contrasts of interest.

The best linear unbiased estimate for the contrast vector is $\hat{\boldsymbol{\gamma}} = B\hat{\boldsymbol{\beta}} = B(X^T X)^{-1} X^T \boldsymbol{M}$ and the variance matrix

$$\mathrm{var}(\hat{\boldsymbol{\gamma}}) = B\mathrm{var}(\hat{\boldsymbol{\beta}})B^T = \sigma^2 B(X^T X)^{-1} B^T. \tag{4}$$

In the present gene expression study, due to scientific interest in cell line effects, interaction and time effects,

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

Now we can define the multi-objective function, which is based on minimisation of the variances of the elements of $\gamma$, that is the diagonal elements $c_i$, taking into

account that subsets of elements of $\gamma$ are considered to be of equal interest, usually because of symmetries in the model. As in [11], the objectives are $t_A = c_\alpha + c_{\alpha+\alpha\beta}$, $t_B = c_\beta + c_{\beta+\alpha\beta}$ and $t_{AB} = c_{\alpha\beta}$ which are penalised based on the constraints $c_\alpha = c_{\alpha+\alpha\beta}$ and $c_\beta = c_{\beta+\alpha\beta}$ which arise from scientific interest. That is, the change in expression between the wild type and mutant cell line with no time delay is of equal interest to the change in expression between the wild type and mutant cell line with 24 hours delay. It is also of equal scientific interest as to how each of the cell lines behaves over time. A penalty is used because enforcing constraints exactly can be unduly restrictive. For example, there may be situations where none of the Pareto optimal designs satisfy the constraints. In extreme cases, it may happen that there are no designs at all that satisfy the constraints.

Based on the constraints, an appropriate penalty function is $D = (c_\alpha - c_{\alpha+\alpha\beta})^2 + (c_\beta - c_{\beta+\alpha\beta})^2$.

Then the penalised objectives are given by $t_i^{(D)} = (1-w)t_i + wD$ where $w \in [0,1)$ is the weight associated with the penalty and $t_i$ is the $i$-th objective.

The Pareto optimal set is then generated for the penalized objectives $t_A^{(D)}$, $t_B^{(D)}$ and $t_{AB}^{(D)}$ corresponding to the effects for cell line, time and interaction respectively. Although those effects are all of interest, interaction is of primary interest in this gene expression study.

[11] generated Pareto optimal designs for gene expression studies with relatively small numbers of available microarray slides using an exhaustive search of the design space. It was also found that a relatively high penalty weight, close to 1, tended to work well.

In practice, for large gene expression studies, such as those for which a large number of slides are to be utilised, it is not feasible to examine all possible designs in an exhaustive search for Pareto optimal designs. In this paper, we adapt the multiple objective metaheuristic method of Pareto simulated annealing to the design of experiments. The metaheuristic method of Pareto simulated annealing was introduced in [4] and [7] to solve multiple objective combinatorial optimisation problems based on employing the criterion of Pareto optimality. Pareto simulated annealing also extends some principles applied in single objective simulated annealing (see [9] for an introduction to simulated annealing).

Our adaptation of Pareto simulated annealing is aimed at finding optimal or near-optimal designs for gene expression studies for situations where it is not feasible to carry out an enumerative search. This is based on employing the criterion of Pareto optimality which is appropriate, as highlighted earlier, since the efficiency of effects of scientific interest are prioritised and are of interest separately. At each iteration in the search based on Pareto simulated annealing, a sample of generating designs explores the design space in an efficient way according to the Pareto simulated annealing algorithm we present. This involves the setting of a number of Pareto simulated annealing parameters as well as the development of appropriate quality measures to assess their performance. In addition, we present algorithms we have developed to search systematically for the optimal values of tuning parameters associated with Pareto simulated annealing based on the incorporation of response surface methodology.

Our algorithms will be demonstrated in the context of a gene expression study

where the objectives in the optimisation problem, corresponding to the effects of particular scientific interest, are

- $t_A^{(D)}$ for cell line

- $t_B^{(D)}$ for time and

- $t_{AB}^{(D)}$ for interaction, where interaction is of primary interest

(with background details provided earlier in this section). Our consideration will include catering for the case where an exhaustive search of the design space for Pareto optimal designs is infeasible due to allowing for a large number of microarray slides to be used in the experiment.

## 2 Pareto simulated annealing strategy

We present our adaptation of the metaheuristic method of Pareto simulated annealing given in [4] and [7] to the design of experiments as follows.

### 2.1 Strategy

The aim of Pareto simulated annealing is to find a good approximation to the exact set of Pareto optimal designs in a relatively short time. In order to do so, a guided search is carried out in the design space using a sample of generating designs. At each iteration of the search, a given generating design is compared to a nearby, or neighbourhood, design. At the beginning of the search, the generating designs move most freely around the design space. The search becomes gradually more selective over time in terms of a reduction in the probability that a generating design will move to a design that provides no improvement in any of the variance objective functions. In addition to the use of generating designs, a set of designs, called the *potentially Pareto optimal set*, is maintained during the search.

Another aspect of the guided search is the information exchange that takes place among generating designs. During the search, generating designs are compared to each other and are influenced to repel each other to be dispersed throughout the variance objective function space.

### 2.2 Strategic concepts

#### 2.2.1 Generating designs

The generating designs are a set of designs used to explore the design space. During the search, individual generating designs may be replaced by neighbouring designs.

## 2.2.2   Neighbourhood

A neighbourhood structure is defined such that it allows for the possibility of moving from a given design to any other design over time. In the case of a gene expression study, the neighbourhood of **d** can be defined to be any design obtained by removing a slide from one configuration and re-allocating it to a different configuration.

## 2.2.3   Multiple objective function

Pareto simulated annealing is concerned with optimising for multiple objectives since the variance objective function is multi-dimensional. For example, for a generating design **d**, the corresponding variance objective function, $f(\mathbf{d})$, is p-dimensional when optimising for $p$ objectives.

## 2.2.4   Potentially Pareto optimal set

The potentially Pareto optimal set is the set of designs obtained by carrying out Pareto simulated annealing. It is maintained during the search so that it is updated at each iteration, that is, each time a new design is visited.

## 2.2.5   Acceptance probability

The set of generating designs consists of one or more individual designs. Moves are made by individual generating designs one at a time during the search. The acceptance probability is a measure of the probability of moving from the current generating design **d** to a randomly selected neighbourhood design **y**. Taking into account multiple objectives, given by variance objective functions, the following two acceptance probability functions are given based on [4].

$$P(\mathbf{d}, \mathbf{y}, T, \Lambda) = \min\{1, \exp(\max_i\{\lambda_i(f_i(\mathbf{d}) - f_i(\mathbf{y}))/T\})\} \tag{5}$$

and

$$P(\mathbf{d}, \mathbf{y}, T, \Lambda) = \min\{1, \exp(\sum_{i=1}^{p} \lambda_i(f_i(\mathbf{d}) - f_i(\mathbf{y}))/T)\} \tag{6}$$

where $T$ is the temperature and $\Lambda = \{\lambda_i\}$ is the set of weights used to influence the dispersion of generating designs.

For each of the acceptance probability functions, given above, the following holds. If a neighbourhood design does at least as well as the generating design for all variance objective functions, then the generating design will be replaced by the neighbourhood design. If the neighbourhood design is less efficient for all objectives, then the probability of moving to the neighbourhood design will always be less than one. In all other cases, the probability of moving to a neighbourhood design:

- is 1 when using the acceptance probability function given by Equation (5) since there is at least one objective that the neighbourhood design is doing better on than the generating design, and

- depends on the weighted sum of differences for each variance objective function when using the acceptance probability function given by Equation (6).

The acceptance probability also depends on the temperature parameter. During the search, the temperature parameter progressively decreases according to a cooling schedule.

### 2.2.6   Cooling schedule

In order to achieve the Pareto simulated annealing strategy described in Section 2.1, the temperature parameter is progressively decreased according to a cooling schedule as follows.

- Select an initial temperature $T_0$, number of steps at each temperature, $l$, and cooling rate $\kappa$ (where $0 < \kappa < 1$)

- For each temperature, $T_k$, do the following

  - Maintain the temperature for $l$ steps
  - After $l$ steps, decrease the temperature. For example, using cooling rate $\kappa$,

$$T_k = \kappa^k T_0 \tag{7}$$

  - Stop when the final temperature is reached

### 2.2.7   Stopping rule

The stopping rule determines when the search ends. One example is to stop after a certain number of designs have been visited. Another example is to stop when the proportion of accepted moves is less than $\epsilon$ % at a given temperature.

### 2.3   Role of weights

The role of the weights, $\Lambda$, is to disperse the generating designs throughout the variance objective space. This is implemented by adjusting the weights used in the acceptance probability function based on information exchange between generating designs so that they act to repel each other in the variance objective space. Firstly, for each generating design, the Euclidean distance is used to determine which of the other generating designs is closest to it as follows. Consider a given generating

design $\mathbf{d}^i$ with corresponding variance objective function $f(\mathbf{d^i})$. Consider the $j$-th generating design $\mathbf{d}^j$ with corresponding variance objective function $f(\mathbf{d}^j)$ in the set of generating designs such that $i \neq j$. The distance between $\mathbf{d}^i$ and $\mathbf{d}^j$ is taken to be $S = \sum_{k=1}^{p} (f(\mathbf{d}^i)_k - f(\mathbf{d}^j)_k)^2$. The closest generating design to $\mathbf{d}^i$ is taken to be the generating design $\mathbf{d}^j$ for which $S$ is minimal and which is not dominated by $\mathbf{d}^i$. Then the weights for each objective, to be applied to the acceptance probability function, are adjusted such that

- for the objectives that the given generating design is doing well on compared to the closest generating design, the weights are increased so that the given generating design is more likely to make a move that continues to improve on those objectives, and

- for the objectives that the given generating design is not doing well on compared to the closest generating design, the weights are decreased so that the given generating design is less likely to be geared towards making a move that improves on those objectives.

If the given generating design $\mathbf{d}^i$ dominates all other designs in the set of generating designs, a closest generating design does not exist and the weights of the given generating design are randomly adjusted. That is, for each of the objectives, each of the associated weights are adjusted separately to increase or decrease with equal probability.

## 2.4   Core Pareto simulated annealing algorithm

**Initialisation**

1. Select an initial set of generating designs $\{\mathbf{d}\}$ at random.

2. For each generating design $\mathbf{d}$, generate an initial weight vector such that $\lambda_i^{\mathbf{d}} = 1/p$ where $p$ is the number of variance objective functions.

3. Initialise the set of potentially Pareto optimal designs, $PP$, to be those designs that are Pareto optimal among the initial set of generating designs.

4. Set the initial temperature such that $T = T_0$.

**Iterative Steps**

For each generating design, $\mathbf{d}$, do the following until the stopping condition is fulfilled.

1. Construct a neighbourhood design $\mathbf{y}$ as follows. Randomly select a configuration in $\{\mathbf{d}\}$ that has at least one slide allocated to it. Remove one slide from that configuration and re-allocate it to a different configuration that is selected at random.

2. If $\mathbf{y}$ is not dominated by $\mathbf{d}$, update the set PP with $\mathbf{y}$.

3. Select the closest generating design, in the objective space, $\mathbf{d^o}$ that is not dominated with respect to $\mathbf{d}$. If there is no such generating design, adjust the weights such that $\lambda_i^{\mathbf{d}} = \lambda_i^{\mathbf{d}}\alpha$ or $\lambda_i^{\mathbf{d}} = \lambda_i^{\mathbf{d}}/\alpha$, each with probability equal to 0.5. Otherwise adjust the weights such that

$$\lambda_i^{\mathbf{d}} = \begin{cases} \lambda_i^{\mathbf{d}}\alpha, & \text{if } f_i(\mathbf{d}) < f_i(\mathbf{d^o}) \\ \lambda_i^{\mathbf{d}}/\alpha, & \text{if } f_i(\mathbf{d}) \geq f_i(\mathbf{d^o}) \end{cases}$$

($\alpha$ is greater than, but close to, 1.)

4. Normalize the weights such that $\sum\limits_i \lambda_i^{\mathbf{d}} = 1$.

5. Accept the neighbourhood design, to replace the generating design, with probability $P(\mathbf{d}, \mathbf{y}, T, \Lambda^{\mathbf{d}})$.

6. If the condition for changing the temperature is fulfilled, decrease the temperature $T$ such that it becomes $\kappa T$, $0 < \kappa < 1$.

## 3   Quality measures

To evaluate the performance of Pareto simulated annealing, the following quality measures are proposed.

### 3.1   Comparison with exact set of Pareto optimal designs

The quality measures presented in this section arise from considering Czyzak and Jaszkiewicz (1998). It is assumed that the exact set of Pareto optimal designs is given.

#### 3.1.1   Number of designs missed

Quality measure $Q_m$ is based on the number of designs that appear in the exact set $R$ of Pareto optimal designs but not in the potentially Pareto optimal set $PP$ obtained by applying Pareto simulated annealing. This constitutes the number of designs missed given by

$$Q_m = \text{card}\{R\} - (\text{card}\{PP \cap R\}). \tag{8}$$

A similar measure is the proportion of designs missed whereby

$$Q_p = \frac{\text{card}\{R\} - (\text{card}\{PP \cap R\})}{\text{card}\{R\}}. \tag{9}$$

In practice, it is convenient to use the empirical logit whereby

$$Q_l = \log{(Q_m + 0.5)}/(\text{card}\{R\} + 0.5 - Q_m) \tag{10}$$

### 3.1.2  Average distance

Quality measure $Q_a$ is based on the average distance of designs in the exact set to the closest design in the set of potentially Pareto optimal designs as follows. For a given design, say $\mathbf{v}$, in the exact set, the closeness of a design $\mathbf{u}$ in the set of potentially Pareto optimal designs is given by

$$c(\mathbf{u}, \mathbf{v}) = \max_{i=1,2,\dots,p} \{0, w_i(f_i(\mathbf{u}) - f_i(\mathbf{v}))\} \tag{11}$$

The weight $w_i = 1/\Delta_i$ where $\Delta_i$ is the range for the $i$-th objective in the exact set. The closest design is that $\mathbf{u}$ for which $c(\mathbf{u}, \mathbf{v})$ is minimized. Following such measurements for all designs in the exact set, $Q_a$ is the average of the distances such that

$$Q_a = \frac{1}{\text{card}\{R\}} \sum_{\mathbf{v} \in R} \{\min_{\mathbf{u} \in PP} \{c(\mathbf{u}, \mathbf{v})\}\}. \tag{12}$$

### 3.1.3  Worst case

Quality measure $Q_w$ presents the worst case scenario as follows. For each design in the exact set, the closest design in the set of potentially Pareto optimal designs is found using Equation (11). After all designs are considered, the worst case is returned such that

$$Q_w = \max_{\mathbf{v} \in R} \{\min_{\mathbf{u} \in PP} \{c(\mathbf{u}, \mathbf{v})\}\}. \tag{13}$$

## 4   Tuning parameters for Pareto simulated annealing

### 4.1   Introduction

The core Pareto simulated annealing algorithm, presented in Section 2.4, involves a number of tuning parameters that affect its performance. In this section, algorithms to search systematically for the optimal values of the tuning parameters are developed. To achieve this, the performance of the core Pareto simulated annealing algorithm is systematically evaluated with the use of response surface methodology. For further background about response surface methodology, see [2] and [1]. Firstly, we develop a parameter selection algorithm to study the impact of the tuning parameters in simple examples for which the exact Pareto optimal set $R$ is known and can be used to inform the choice of parameters in larger problems. Following this, an adaptive algorithm for

the selection of tuning parameters is developed in the practical case when the exact set is not known.

The tuning parameters of the core algorithm to be considered are summarised as follows.

**Number of generating designs (NG):** this is the number of designs used to explore the design space such that each visits successive designs in its neighbourhood.

**Initial temperature (IT):** this is the temperature that is set at the beginning of the algorithm.

**Cooling rate (CR):** this is the rate, $\kappa \in (0, 1]$, at which the temperature $T$ is decreased such that it becomes $\kappa T$ following the completion of a temperature level.

**Repulsion coefficient (RC):** this is the multiplicative factor, $\alpha \geq 1$, used to adjust weights to allow for the dispersion of generating designs.

**Acceptance rule (AR):** this is the rule used to calculate the probability of moving from the current generating design to a randomly selected neighbourhood design. Rule 0 and rule 1, given by equations (5) and (6) respectively, are considered.

The number of designs, initial temperature and cooling rate are quantitative parameters. The consideration of the repulsion coefficient is to assign two values, one to indicate no repulsion ($\alpha = 1$) and the other to indicate repulsion (choose $\alpha > 1$). The acceptance rule has two values corresponding to the two options proposed for the acceptance probability function.

## 4.2   Central composite experimental plan

In order to find the optimal values for the tuning parameters, a sequence of experiments is conducted. Each such experiment is defined by a central composite design, which were introduced in general in [3]. In the present context, the quantitative parameters, number of generating designs, initial temperature and cooling rate, are each assigned 5 values, corresponding to very low, low, medium, high and very high. The values for each quantitative parameter are typically set so that the interval between the low and medium levels is equal to that for medium and high. Furthermore, these intervals are typically twice the size of that for the interval between very low and low and and that for high and very high.

The central composite design is constructed as follows. Firstly, for the quantitative parameters, form the setting combinations consisting of:

- all combinations of low and high levels for the quantitative parameters,

- the combination corresponding to each quantitative parameter set to the medium level,

- the combinations arising from setting each quantitative parameter in turn to very low while the others are set to medium and

- the combinations arising from setting each quantitative parameter in turn to very high while the others are set to medium.

To consider the repulsion coefficient and acceptance rule also, each of the setting combinations formed for the quantitative parameters is carried out in the absence and presence of repulsion and for both acceptance rules. All setting combinations constitute the central composite design.

A single experiment then consists of a specified number of replicates of the central composite design.

## 4.3 Analysis

The analysis of a single experiment is performed using multiple linear regression. Each tuning parameter is represented by a variable $x$. For the quantitative variables, number of generating designs, initial temperature and cooling rate, the values of $x$ are $-1.5$, $-1$, $0$, $1$ and $1.5$. For the repulsion coefficient and acceptance rule, the values of $x$ are -1 and 1.

The quality measure, $Q$, forms the response variable. Examples of such measures were presented in Section 3.1 for the case where the exact set is known. Later in this section, practical cases where the exact set is not known are catered for.

The linear model is defined to be

$$ML : E(Q) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i$$

and the quadratic model is defined to be

$$MQ : E(Q) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \gamma_i x_i^2 + \sum_{i<j} \gamma_{ij} x_i x_j.$$

In what follows, algorithms for finding optimal values for the tuning parameters are proposed based on the core algorithm and use of response surface methodology.

## 4.4 Parameter selection algorithm

For cases where the exact set $R$ is known, the following algorithm for the selection of suitable values for the Pareto simulated annealing tuning parameters is proposed.

**Initialisation**

1. Set the parameter values for the first experiment.

2. Using the appropriate combinations of parameter values, form the first composite design.

3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the first central composite design.

4. Calculate the quality measure $Q$ for each run.

5. Fit the linear and quadratic models given by $ML$ and $MQ$ respectively.

6. Determine whether the quadratic model provides a substantial improvement over the linear model. A formal test to do so is given as follows. Calculate the F-statistic $\frac{(R^2_{MQ}-R^2_{ML})/(\nu_{ML}-\nu_{MQ})}{(1-R^2_{MQ})/\nu_{MQ}}$. where $\nu_{ML}$ and $\nu_{MQ}$ are the degrees of freedom for $ML$ and $MQ$ respectively. Compare the calculated value with $F_{\nu_{MQ}-\nu_{ML},\nu_{MQ}}$. Based on the critical point of $F$ at the 10% level of significance, accept the quadratic model if the calculated value of $F$ exceeds the critical point.

7. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. A formal specification is to adjust the tuning parameter values in proportion to their estimated coefficients. Thus, if the coefficients of the $k$ PSA tuning parameters, $x_i$, are $\beta_i$ and $x_1$ is changed by $\Delta$ then set

$$x_j = \frac{\beta_j}{\beta_1}\Delta \qquad (14)$$

for $j = 2, ..., k$. The value of $\Delta$ is typically taken so that the move is 1 in terms of the coded units. That is:

$$\Delta = (1 + \beta_1^2 + ... + \beta_k^2)^{-0.5}. \qquad (15)$$

In practice, there may be restrictions on the feasible range of values for the tuning parameters and an approximation to the estimated direction of steepest descent is used. If, however, the quadratic model is accepted, proceed to the optimisation step.

**Iterative Steps**

The following steps are repeated until it is determined that the quadratic model is appropriate.

1. Using the method of steepest descent, set the parameter values for the current experiment.

2. Using the parameter values, form the current central composite design.

3. Perform a number of replicate runs of the core Pareto simulated annealing algorithm for each combination of parameters in the current central composite design.

4. Calculate the quality measure $Q$ for each run.

5. Fit the linear and quadratic models given by $ML$ and $MQ$ respectively.

6. Determine whether the quadratic model provides a substantial improvement over the linear model.

7. If the quadratic model provides no substantial improvement over the linear model, apply the method of steepest descent. If, however, the quadratic model is accepted, proceed to the optimisation step.

**Optimisation step**

Following the determination of an appropriate quadratic model, the optimal values for the parameters can be estimated as follows.

1. Choose $x_1, x_2, \ldots, x_k$ to find the least value of the quadratic function

$$\hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\gamma}_i x_i^2 + \sum_{i<j} \hat{\gamma}_{ij} x_i x_j,$$

within the specified domain, where $\{\hat{\beta}_0, \hat{\beta}_i, \hat{\gamma}_i, \hat{\gamma}_{ij}\}$ are the least squares estimates obtained from the final experiment. If this least value lies on the boundary, it is not a minimum (unless a minimum lies precisely on the boundary). If it is not a minimum, use it to define the direction of steepest descent and return to Iterative Step 7.

2. Calculate the corresponding values for the tuning parameters on the original scale.

For cases where the exact set is known, this algorithm can be applied to obtain insight into the importance and effect of each of the tuning parameters. Moreover, it provides the basis for the development and testing of the adaptive algorithm described in the next section. Applications are presented later in this paper.

## 4.5 Adaptive Pareto simulated annealing algorithm

In practical cases, it may not be feasible to determine the exact set of Pareto optimal designs. An adaptive algorithm that caters for such cases can be implemented in a manner identical to the parameter selection algorithm except that the exact set $R$ is replaced by a suitably constructed reference set $U$ that is defined by cumulatively combining all of the potentially Pareto optimal designs from all experiments as the iterations proceed. This updating provides the opportunity for the set $U$ to continue to improve as a set of potentially Pareto optimal designs as the algorithm progresses. At each iteration, the reference set $U$ is used in place of $R$ for the calculation of quality measures described in Section 3.1.

# 5 Applications

In this section, the parameter selection and adaptive algorithms are applied to the gene expression study, introduced in Section 1, that investigates the genetic basis of leukaemia. The penalised objectives for the effects for cell line, time and interaction are of interest and the penalty weight used is $w = 0.9999$. We will consider the generation of Pareto optimal designs for this gene expression study so that it is considered to be large, which is characterised by having a relatively large number of microarray slides available.

The multiple objective combinatorial optimisation problem can be summarised as follows:

- There are six possible comparisons ($i = 1, \ldots, 6$), any one of which can be made on a single slide.

- There are $n$ slides.

- Let $d_i$ be the number of slides used for comparison $i$.

The objectives to be minimized are:

$$
\begin{aligned}
t_A &= (1-w)(c_\alpha + c_{\alpha+\alpha\beta}) + wD \\
t_B &= (1-w)(c_\beta + c_{\beta+\alpha\beta}) + wD \\
t_{AB} &= (1-w)c_{\alpha\beta} + wD
\end{aligned}
$$

where

$$
D = (c_\alpha - c_{\alpha+\alpha\beta})^2 + (c_\beta - c_{\beta+\alpha\beta})^2.
$$

The constraints are

$$
d_1, \ldots, d_6 \geq 0
$$

and

$$
d_1 + \cdots + d_6 = n.
$$

A simulated annealing algorithm (SSA) is used to tackle this combinatorial problem. The SAA has 5 parameters (NG, IT, CR, RC, AR) and the intermediate objective is to tune (optimise) the performance of the SAA. The tuning objective is defined in terms of Q which is the number of designs missed by a run of the SAA relative to the set U of all potentially Pareto optimal designs obtained at the end of the run (U-Q is therefore the number of designs contributed by the run).
Minimise Q
subject to:

$$
\begin{aligned}
NG &\in \{10, 20, 40, 60, 70\} \\
IT &\in \{0.5, 5, 40, 400, 1200\} \\
CR &\in \{0.4, 0.5, 0.7, 0.9, 1.0\} \\
RC &\in \{0, 1\} \\
AR &\in \{0, 1\}
\end{aligned}
$$

## 5.1 Parameter selection algorithm: 36 slides

As an illustration, consider the case where 36 slides are available. In this situation, there are 749, 398 possible designs. It is feasible to carry out an enumerative search of the design space and the Pareto optimal set obtained consists of 63 designs. This constitutes the reference set.

Now suppose that the parameter selection algorithm is applied. A central composite experimental plan for the Pareto simulated annealing parameters is adopted as follows. The number of generating designs, initial temperature and cooling rate consist of 5 levels each, corresponding to very low, low, medium, high and very high. The values for those parameters are given in Table 1.

For a given acceptance rule, the method of setting the initial temperature in Table 1 is carried out as follows. Let the proportion of moves be the number of times a generating design is replaced by a neighbourhood design divided by the total number of visits at a given temperature. The initial temperatures were chosen to correspond to approximate proportions of moves: 30%, 40%, 60%, 80% and 90%. These proportions depend on acceptance probabilities and vary with generating designs. The correspondences between initial temperature and proportion of moves were estimated from pilot runs. Note that this was the method used for setting initial temperatures in all applications presented in this paper.

The number of steps is fixed to be 600. The values for the repulsion coefficient and acceptance rule used each consist of 2 levels; (1, 1.05) and (rule 0, rule 1). Note that rule 0 and rule 1 are given by Equations (5) and (6) respectively. Thus the experimental plan consists of 15 setting combinations to vary the number of generating designs, initial temperature and cooling rate for a given level of the repulsion coefficient and acceptance rule.

The 15 combinations are carried out for each combination of the repulsion coefficient and acceptance rule, with a total of 60 settings. The number of replicates for each setting is 4 thus there are 240 runs in total. The stopping criterion is the completion of the temperature level that results in having visited at least 40, 000 designs, which is 5.333% of the design space.

**Table 1**: Values for parameters that have 5 levels for the first experimental plan for the gene expression study with 36 slides

| NG | 1 | 4 | 10 | 16 | 19 |
|---|---|---|---|---|---|
| Initial T rule 0 | 0.00002 | 0.00003 | 0.000085 | 0.0004 | 0.002 |
| Initial T rule 1 | 0.00005 | 0.00009 | 0.00035 | 0.0013 | 0.007 |
| CR | 0.4 | 0.5 | 0.7 | 0.9 | 1 |

The quality measure used is based on the number of Pareto optimal designs, $M$, in the exact set missed by the application of the particular Pareto simulated annealing setting.

In particular, the empirical logit, introduced in Section 3.1, is used such that

$$Y = \log{(M + 0.5)}/(63.5 - M) \qquad (16)$$

given that there are 63 designs in the exact set.

The results from the first Pareto simulated annealing experimental plan were analysed using linear models in $R$, [10].

The adjusted R-squared for the linear model was found to be 34% compared to 60% for the quadratic model. The linear model is statistically significant, inasmuch as there is overwhelming evidence against a null hypothesis that all the coefficients except the intercept are 0, but the quadratic model is a statistically significant improvement. The F-ratio for testing the null hypothesis that the coefficients of all the quadratic and cross-product terms are 0 is 12.67. When compared with the quantiles of an F-distribution with 13 and 221 degrees of freedom, for the numerator and denominator respectively, there is evidence to reject this null hypothesis at the $0.6 \times 10^{-5}$ level. However, the least value of the fitted quadratic model, subject to the constraints, occurs on a boundary. In the coded units, the least value of the empirical logit is $-1.941$ when the number of generating designs, initial temperature, cooling rate, repulsion coefficient and acceptance rule are set at 0.13, $-1.5$, 1.5, $-1$ and 1 respectively. The linear model suggests that an increase in the number of generating designs and the cooling rate and a reduction in the initial temperature provide improvements.

Taking the estimates and feasible values for the coefficients into account, the parameter settings from the first experimental plan are modified to the values given in Table 2. This forms the second central composite experimental plan. Note that the initial temperatures were chosen to correspond to approximate proportions of moves: 10%, 20%, 40%, 60% and 70%.

**Table 2**: Values for parameters that have 5 levels after applying steepest descent for the gene expression study with 36 slides

| NG | 10 | 13 | 19 | 25 | 28 |
|---|---|---|---|---|---|
| Initial T rule 1 | 0.00001 | 0.00003 | 0.00009 | 0.00035 | 0.0007 |
| CR | 0.7 | 0.75 | 0.85 | 0.95 | 1 |

In addition, it is found that the model from the first experimental plan suggests that the low level of repulsion, which corresponds to no repulsion, is better and that acceptance rule 1 is better. Thus the second experimental plan consists of the 15 setting combinations from the modified parameter settings for the number of generating designs, initial temperature and cooling rate and applies no repulsion and uses acceptance rule 1. Furthermore, to maintain a total of 240 runs as carried out in the first plan, each setting in the second plan has 16 replicates. The stopping rule and number of steps from the first plan are preserved.

The results from the second plan were analysed in $R$. The linear model is found to be only just statistically significant and has an adjusted R-squared of 3%. Moving to the quadratic model, the adjusted R-squared increases to 63% and it is statistically

significant. The quadratic model implies being near the minimum of a quadratic surface. Thus no further iterations are required and the quadratic model is used to find the co-ordinates of the minimum.

Using Solver in Excel, the co-ordinates of the minimum of the quadratic surface were found and the values for each of the parameters are given in Table 3.

**Table 3**: Estimated optimal parameter values for the empirical logit quality measure for the gene expression study with 36 slides

| Parameter | Estimated value |
|-----------|-----------------|
| NG        | 28              |
| Initial T | 0.00005         |
| CR        | 1               |

The logit response and associated analysis presented is appropriate for the current application. In addition, recall that the average distance and worst case quality measures were introduced in Section 3.1 as alternatives. Using the second experimental plan for the Pareto simulated annealing settings, the current application, that is, the second experimental plan, is analysed in R with those alternative quality measures. For the average distance quality measure, the optimal values for the tuning parameters were found to be NG = 28, Initial T = 0.00015 and CR = 1. For the worst case distance quality measure, the optimal values for the tuning parameters were found to be NG = 10, Initial T = 0.00018 and CR = 1.

## 5.2   Adaptive Pareto simulated annealing algorithm

In what follows, we consider applying the adaptive algorithm for the gene expression study with 160 slides available. When 160 slides are available, It is infeasible time-wise to find the set of all Pareto optimal designs therefore the adaptive approach is necessary. Firstly, we adopt a central composite experimental plan arising from the parameter values given in Table 4. The initial temperatures were chosen to correspond to approximate proportions of moves: 42.5%, 50%, 65%, 80% and 87.5%.

**Table 4**: Values for parameters that have 5 levels for the experimental plan for the gene expression study with 160 slides

| NG        | 10  | 20 | 40  | 60  | 70   |
|-----------|-----|----|-----|-----|------|
| Initial T | 0.5 | 5  | 40  | 400 | 1200 |
| CR        | 0.4 | 0.5| 0.7 | 0.9 | 1    |

The central composite experimental plan consists of 15 setting combinations. Each setting combination for the central composite experimental plan was replicated 4

times, with a total of 60 runs. The number of steps is fixed to be 60 and the stopping criterion for each run is the completion of the temperature level that results in having visited at least $100,000$ designs. The quality measure used is the empirical logit of the number of designs missed by the application of the particular Pareto simulated annealing setting, taking the reference set to be $U$.

After the first central composite experimental plan was executed, the reference set $U$ was constructed to be the designs deemed to be Pareto optimal among the 60 potentially Pareto optimal sets obtained from the runs. For this experiment, $U$ was found to consist of 940 designs. The results were analysed in $R$.

The linear model is not found to be statistically significant (P=0.22). For the quadratic model, the adjusted $R^2$ is 54%. In particular, note that all three interaction terms are statistically highly significant.

The quadratic model implies being near the minimum of a quadratic surface. Thus no further iterations are required and the optimisation step is carried out as follows.

The quadratic model is used to find the co-ordinates of the minimum using Solver in Excel. The optimum values of the Pareto simulated annealing parameters within the region explored were found to be 70 for the number of generating designs, 0.5 for initial temperature and 0.6211 for the cooling rate.

The algorithm was carried out from the optimum values for 60 runs. Then $U$ was updated from the set of 940 designs to find those Pareto optimal among those 940 designs as well as the 60 potentially Pereto optimal sets from the runs at the optimum. This leads to an updated $U$ consisting of a set of 924 designs obtained by adding 172 designs and removing 188 designs. Of the 172 designs added, 156 of those designs satisfy the constraints for the subsets of interest which were presented in Section 1.

The updated set $U$ consists of 50 designs that provide minimum variance for interaction of 0.025 and it is interaction that is of primary interest in this gene expression study. Each of the 50 designs are also those that satisfy the constraints. During the updating of $U$, due to carrying out 60 runs at the optimum, 10 of those 50 designs were added, including a design that provides equal variance for cell line and time effects of 0.01875 in addition to having minimum variance for interaction among the designs in $U$.

The improvement obtained in the optimisation step can also be seen in Figure 2 that shows the relative efficiencies for the three objectives $t^A$, $t^B$ and $t^{AB}$ for each of the Pareto optimal designs at the first adaptive step and then after the final optimisation. Note that in both cases, the relative efficiencies are with respect to the final reference set, $U$. This comparison shows an obvious improvement but the gaps in suggest that a number of Pareto optimal designs have still been missed.
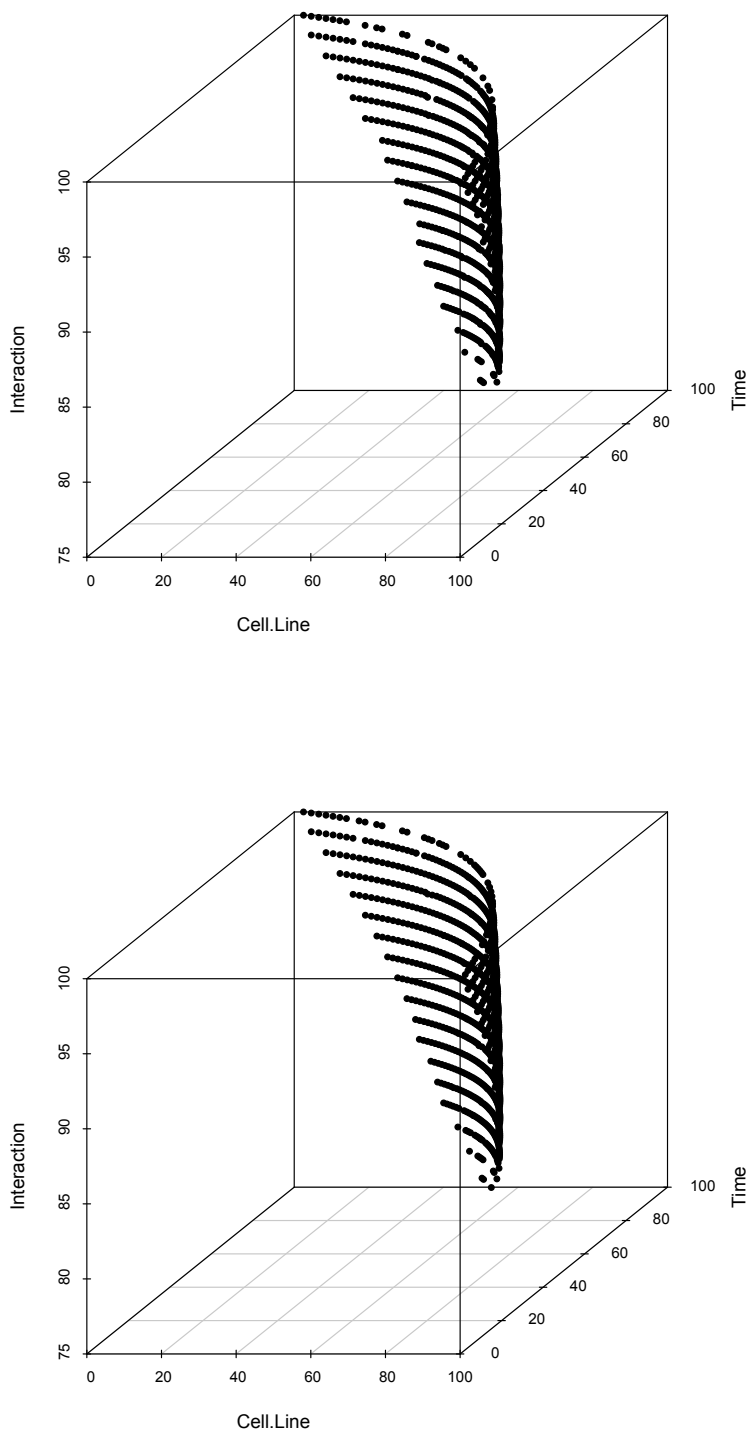
**Figure 2**: Relative efficiencies for $2 \times 2$ experiment with 160 slides, at the first step (upper panel) and after the final optimisation (lower panel).

Now consider fitting models for the first Pareto simulated annealing plan for 160 slides using the average and worst case quality measures that were defined in Section 3.1 except that $U$ is the reference set as described in Section 4.5. The analysis for both measures was carried out in R.

When either the average distance or worst case quality measure is used in place of the logit response, the optimum values of the Pareto simulated annealing parameters for the number of generating designs and the initial temperature remain at the upper end and lower end of their ranges, 70 and 0.5 respectively. The optimum cooling rate is 0.64 for the average distance and 0.70 for the worst case quality measures, relatively near 0.62 found for the logit response quality measure. In this application, at least, the choice of quality measure has little effect on the optimum values of the Pareto simulated annealing parameters.

# 6    Conclusion

This paper demonstrates how the metaheuristic method of Pareto simulated annealing can be adapted to the context of the design of experiments. Furthermore we demonstrate how to incorporate response surface methodology to find values for the tuning parameters and the use of quality measures to assess its performance. We have successfully applied our algorithms to generate optimal and near-optimal designs in the case of a gene expression study and provide a potential platform for other types of applications in experimental design to be explored.

# References

[1] G. E. P. Box and N. R Draper. *Empirical Model Building and Response Surfaces.* New York: J. Wiley & Sons., 1987.

[2] G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters.* New York: J. Wiley & Sons., 1978.

[3] G. E. P. Box and K. B. Wilson. On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society: Series B*, 13(1):1–45, 1951.

[4] P. Czyzak and A. Jaszkiewicz. Pareto simulated annealing-a metaheuristic technique for multiple objective combinatorial optimisation. *Journal of Multi-Criteria Decision Analysis*, 7:34–47, 1998.

[5] G. F. V. Glonek and P. J. Solomon. Factorial and time course designs for cDNA microarray experiments. *Biostatistics*, 5:89–111, 2004.

[6] P. A. Gregory, A. G. Bert, E. L. Paterson, S. C. Barry, A. Tsykin, G. Farshid, M. A. Vadas, Y. Khew-Goodall, and G. J. Goodall. The mir-200 family and mir-

205 regulate epithelial to mesenchymal transition by targeting zeb1 and sip1. *Nature cell biology*, 10(5):593–601, 2008.

[7] A. Jaszkiewicz. *Multiple Objective Metaheuristic Algorithms for Combinatorial Optimization*. PhD thesis, Poznan University of Technology, Poznan, 2001.

[8] D. Nguyen, A. Arpat, N. Wang, and R. Carroll. DNA microarray experiments: biological and technical aspects. *Biometrics*, 58:701–717, 2002.

[9] M. Pirlot. General local search methods. *European Journal of Operational Research*, 92:493–511, 1996.

[10] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. ISBN 3-900051-07-0.

[11] P. S. Sanchez and G. F. V. Glonek. Optimal designs for two-colour microarray experiments. *Biostatistics*, 10(3):561–74, 2009.

[12] S. Searle. *Linear Models*. New York: J. Wiley & Sons., 1971.

[13] G. K. Smyth, Y. H. Yang, and T. Speed. Statistical issues in cDNA microarray data analysis. *Methods in Molecular Biology*, 224:111–136, 2003.

[14] Y. H. Yang and T. P. Speed. *Design and analysis of comparative microarray experiments. In T. P. Speed (Ed.), Statistical analysis of gene expression microarray data*. Boca Raton: CRC Press., 2003.