Nonparametric and semiparametric response surface methodology: a review of designs, models and optimization techniques

Laura Villanova[1] Rob J Hyndman[2] Kate Smith-Miles[1] Irene Poli[3]

29 October 2013

Abstract

Since the introduction of Response Surface Methodology in the 1950s, there have been many developments with the aim of expanding the range of applications of the methodology. Various new design, modeling and optimization techniques have been introduced for coping with unknown input-output relationships, costly or time-consuming experimental studies and irregular experimental regions (e.g., non-cubic or non-spherical regions induced by constraints in the input variables). Such developments may involve many different research areas simultaneously (e.g., the statistical design of the experiments, multivariate modeling, and multi-objective optimization). Experts in various research fields have been embracing and combining methodologies from other areas in order to achieve an improved efficacy and efficiency. This article aims to throw light on these methodological developments and provide a self-contained literature review with a statistical perspective.

Key words: Response surface methodology; experimental design; non-linear models; global optimization; multi-objective optimization.

1 Introduction

Response Surface Methodology (RSM) is a sequential procedure that combines techniques of statistical design and numerical optimization in order to optimize processes and product designs (Montgomery, 2009; Myers et al., 2004). As was stated by Mead and Pike (1975), many of the RSM fundamental ideas have been being used and discussed since the 1930s; however, the actual method is usually considered as having originated in the 1950s with the work by Box and Wilson (1951). Since then, RSM has been a fertile field of research, and its development has been documented extensively by the statistical community (Hill and Hunter, 1966; Mead and Pike, 1975; Myers et al., 1989; Montgomery, 1999; Myers et al., 2004; Robinson et al., 2004; Khuri and

[1]School of Mathematical Sciences, Monash University, Clayton, Australia.
[2]Department of Econometrics & Business Statistics, Monash University, Clayton, Australia.
Email: rob.hyndman@monash.edu
[3]Department of Environment Science, Informatics and Statistics, University Ca' Foscari of Venice, Venice, Italy.

Mukhopadhyay, 2010; Dellino et al., 2010). A parallel body of research has also been developing within the computer science community; we refer to the research on model-based optimization, which has been growing since the early 2000s (Simpson et al., 2001, 2004; Jin, 2005; Kleijnen et al., 2005; Chen et al., 2006; Wang and Shan, 2007; Simpson et al., 2008; Kleijnen, 2009; Jin, 2011).

On the one hand, statistical survey papers have provided detailed and exhaustive reviews of standard RSM designs, models and optimization procedures. On the other hand, computing survey papers have focused on procedures for deterministic simulations; that is, simulations which differ from physical experimentation in that there is no random error associated with the output, so that the same response values are obtained when running the code with the same inputs (Sacks et al., 1989). Often, in the context of physical experimentation, there are situations in which the standard RSM is not applicable, and deterministic methodologies cannot account for the random variation that appears in real systems. Examples of such cases include scientific and industrial settings where the aim is to optimize one or more responses globally, subject to a limited budget of implementable experiments and/or constraints on the many design variables. Over recent years, the challenges of such settings have been attracting the interest of researchers from many different fields and nurturing the development of new approaches, methodologies and algorithms (Knowles and Hughes, 2005; Villanova et al., 2010; Carta et al., 2011; Kleijnen et al., 2005; Jones et al., 1998). The aim of this survey paper is to provide a self-contained review of nonstandard RSMs and model-based optimization in the presence of random variation.

The paper is organized as follows. Section 2 will summarize the standard RSM approach and will indicate the need for Nonparametric and Semiparametric RSMs. Section 3 will review the procedures that have been developed and used in the context of Nonparametric and Semiparametric RSMs; in particular, the section will provide an overview of designs, models and optimization methodologies. Finally, Section 4 will conclude the paper with a brief discussion of future directions and concluding remarks.

2 Response surface methodology and beyond

The basic assumption in RSM is that a response Y is a function of a set of design variables $X_1, X_2, \ldots, X_p$. The methodology aims to determine the levels of the input variables that maximize the response (Balkin and Lin, 2000). This is achieved by examining the search space locally, while leading the experimenter efficiently to the general vicinity of the optimum (Montgomery, 2009). Typically, experimental design strategies (such as 2^k factorial and fractional factorial designs, the central composite design, the small composite design, and the Box-Behnken design) are used to collect a small set of observations efficiently within a restricted area of the search space.

A simple polynomial model is then fitted to the collected data (Montgomery, 2009). Generally, a common error variance σ^2 is assumed, and the coefficients are estimated by ordinary least squares (Myers et al., 1989). Despite their simplicity, polynomial models generally provide a reasonable local approximation of the true response function, because the interest is in a restricted area of the search space (Montgomery, 2009). Typically, a first-order model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{1}$$

is fitted during the initial stages of the procedure. Indeed, the region under investigation is likely to be remote from the optimum, and as a consequence, a model featuring little or no curvature is expected to provide a good fit. Subsequently, the search is directed towards the most promising area of the operability region. New experiments are then performed along the direction in which the fitted response increases (decreases) most rapidly; such a direction is referred to as the path of steepest ascent (descent). This procedure is iterated until the lack of fit of a first-order model is detected; this indicates the presence of curvature in the system, and thus the vicinity of the optimum condition. Once the region of the optimum has been found, a second-order model

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{i=1}^{p} \beta_{ii} X_i^2 + \sum_{i<j} \sum_{i<j} \beta_{ij} X_i X_j + \varepsilon \tag{2}$$

is fitted, and an analysis is performed in order to locate the optimum (Montgomery, 2009).

There are some situations in which the response cannot be accommodated by polynomial models (Myers et al., 2004; Myers, 1999). This may occur when the investigation encompasses a large portion of the search space, possibly even extending over the entire region of operability (Myers et al., 2004; Montgomery, 2009). In such a context, it is likely that non-linearity and complexity will arise, and therefore, suitable modeling and optimization techniques must be adopted. Nonparametric and semiparametric RSMs (NSRSM) provide such techniques. In particular, nonparametric RSM makes use of kernel-based regression, because the primary focus of the analysis is on prediction (Myers et al., 2004; Myers, 1999). On the other hand, semiparametric RSM employs a specific model that is capable of providing an interpretative function (Myers et al., 2004; Myers, 1999).

Nonparametric and semiparametric RSMs have much in common with the Design and Analysis of Computer Experiments (DACE) methodology. DACE is a statistical methodology that was introduced by Sacks et al. (1989). It consists of using the data from a given experimental design to build a model that approximates the system's behavior. Both stochastic and deterministic models are used. Experimentation with a stochastic simulation model is much like experimentation with a physical system (Myers et al., 2004). In contrast, experimentation with a deterministic

computer model differs from physical experimentation by the absence of random error (Sacks et al., 1989). The model originally used with DACE is referred to as the deterministic Gaussian stochastic process model (GP model), also known as kriging in the geostatistics and other spatial statistics literature. In addition to providing a good fit and an efficient predictor of the response to untried inputs, the GP model also allows for estimates of the predictor uncertainty (Sacks et al., 1989). The estimated predictor uncertainty is used to augment the initial experimental design by means of various design criteria and sequential sampling algorithms (Sacks et al., 1989). In particular, where the predictor uncertainty is high, additional design points are chosen, so that attention is paid to parts of the space that are relatively unexplored (Jones, 2001). DACE originally used the deterministic Gaussian stochastic process model and sequential sampling; other models, such as neural networks and thin plate splines, have been introduced subsequently. DACE has been employed extensively in the engineering field for dealing with expensive design optimization problems. In such contexts, both physical experimentation and computer simulations can be impractical. Specifically, physical experimentations can be too time consuming, too expensive, or even impossible to carry out, while simulations can be infeasible due to their excessively high computational costs (Sacks et al., 1989). Typical examples of some expensive optimization problems to which the DACE methodology has been applied include computational fluid dynamics, the study of electronic circuits, and finite element analysis of the stress on a mechanical product (Bates et al., 1996; Grierson and Pak, 1993; Lee and Hajela, 1996; Büche et al., 2004; Pierrot, 1999; Jin et al., 2002).

While RSM and NSRSM differ in many ways (design, modeling and optimization techniques), the distinction between DACE and NSRSM is less clear. The main difference lies in their 'native' domains of application. DACE was developed as a modeling technique for coping with deterministic, expensive simulations of complex engineering systems (Sacks et al., 1989), whereas NSRSM has been applied primarily in the biopharmaceutical and biotechnological fields (Myers et al., 2004). In what follows, DACE will be considered as belonging to the NSRSM class. A parallel body of research, with roots in DACE and NSRSM, has been developed in the field of automatic algorithm parameter optimization. The main technique that has been developed in this field is Sequential Parameter Optimization (Bartz-Beielstein et al., 2005). Its purpose is to quantify the importance of each parameter, as well as the importance of the interactions between parameters. The interpolation of performances between parameter settings and the extrapolation to previously unseen regions of the parameter space are supported as well. This technique can aid substantially in the understanding or improvement of an algorithm (Hutter et al., 2009, 2010).

3 Nonparametric and semiparametric response surface methodologies

Let $X_1,\ldots,X_p$ be p explanatory variables and let $Y_1,\ldots,Y_q$ be q response variables. Assume that $(\boldsymbol{x}_n^T,\boldsymbol{y}_n^T)$ is a set of training data where $\boldsymbol{x}_n=(x_{n1},\ldots,x_{np})^T$, $n=1,\ldots,N$. Then, denote by $\boldsymbol{X}$ the $N\times p$ matrix of observed inputs with the nth row given by $\boldsymbol{x}_n^T$. Similarly, denote by $\boldsymbol{Y}$ the $N\times q$ matrix of observed outputs with the nth row given by $\boldsymbol{y}_n^T$. In the presence of a single response ($q=1$), the output is denoted by the N-dimensional vector of measured response values $\boldsymbol{y}=(y_1,\ldots,y_N)^T$.

3.1 Experimental designs

3.1.1 Single-response designs

Standard RSM designs such as the full/fractional factorial and central composite designs (Myers et al., 2004; Khuri and Mukhopadhyay, 2010) apply in the presence of a regular design region. A regular design region can be represented as a space that is bounded by a p-dimensional hypercube or hypersphere; any point on or inside the cube or sphere constitutes a candidate design point. However, practical problems might require corners to be chopped off the hypercube, slices to be taken off the hyperspheres, or holes to be dug in the design region (Kennard and Stone, 1969). Furthermore, in the case of expensive or time-consuming experimentation, the number of trials that can be implemented is limited. This number is typically small relative to the number of experiments that are conceptually possible. In such contexts, applying standard RSM designs often leads to excessively large numbers of experimental runs being required (Kennard and Stone, 1969).

If nonstandard situations occur (e.g., an irregular experimental region, or a small sample size), computer-generated designs are preferable (Montgomery, 2009). Computer-generated designs are constructed using computer programs. They are referred to as optimal designs because they are optimal with respect to some particular criterion. The most common optimality criteria are D-optimality, A-optimality and G-optimality (Montgomery, 2009). A- and D-optimality are concerned with the estimation of the regression coefficients (Montgomery, 2009), and involve minimizing a function of the covariance matrix of the least squares estimates of the parameters, namely the determinant (D-optimality or generalized variance) and the trace (A-optimality or average variance) (Santner et al., 2003). G-optimal designs are concerned with the prediction of the response, and aim to minimize the maximum scaled prediction variance over the design region (Montgomery, 2009). The usual approach with computer-generated designs is to specify a model (typically a polynomial), determine the region of interest, select the number of runs to perform, and specify the optimality criterion. An algorithm is then employed to select the design points

from a set of candidate points; the candidate points are typically a grid of points spaced over the feasible design region (Montgomery, 2009).

With the advent of DACE, various other criteria-based designs have been developed. They are more difficult to implement than optimal designs. Indeed, a Gaussian stochastic process model is usually assumed, and the design criteria are functions of its unknown parameters (Santner et al., 2003). Furthermore, analytical results are difficult to obtain and have been found in only a few special cases (Santner et al., 2003). For a fixed number of runs and for a specific correlation structure, these design criteria aim to choose a design that will predict the response well for untried inputs in the experimental region (Sacks et al., 1989). Functions of the mean squared prediction error (MSPE), such as the Integrated Mean Squared Error (IMSPE) and the Maximum Mean Squared Error (MMSPE), are considered (Sacks et al., 1989). Given an experimental design $D = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_S)$ in the experimental region E, and the data $\boldsymbol{y} = (f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_S))$, let $\hat{y}(\boldsymbol{x})$ be the predictor of $y(\boldsymbol{x})$ at an untried value of $\boldsymbol{x}$. The IMSPE criterion chooses the design D to minimize $\int_E MSPE[\hat{y}(\boldsymbol{x})]\phi(\boldsymbol{x})d\boldsymbol{x}$ for a given weight function $\phi(\boldsymbol{x})$, while the MMSPE criterion chooses the design D to minimize $\max_{\boldsymbol{x}\in E} MSPE[\hat{y}(\boldsymbol{x})]$. The IMSPE and MMSPE design criteria are generalizations of the classic A-optimality and G-optimality (Santner et al., 2003). A third criterion is the maximum entropy. The entropy is the opposite of the information (Shannon Information) and the maximum entropy criterion aims to maximize the information I about a parameter θ which is contained in a statistical model. The parameter is believed to describe a measured response, and corresponds to the vector of correlation parameters in the Gaussian process model (Santner et al., 2003). Knowledge about θ is specified by a probability density or mass function $[\theta]$, which could be either a prior distribution or a posterior distribution. When the design space is discrete, the maximum entropy criterion is as follows. Let $[\theta|D]$ denote the posterior distribution of θ, given the data obtained using the design D. The amount of information in $[\theta]$ about θ before the experiment is $I = \int[\theta]\ln(\theta)d\theta$, and the amount of information about θ after the experiment is $I_D = \int[\theta|D]\ln(\theta|D)d\theta$. When evaluating a design, the expected change in the information from the prior distribution $[\theta]$ to the posterior distribution $[\theta|D]$ is considered. The design D is chosen to maximize the entropy of the joint distribution of the responses given the design (Fang et al., 2006; Santner et al., 2003).

Along with criteria-based designs, designs that are capable of filling up the factor space in a uniform fashion have also been developed (Kennard and Stone, 1969; Bates et al., 1996). These are space-filling designs. Because the primary concern of NSRSM is predictive accuracy (Santner et al., 2003), space-filling designs are of particular interest. Indeed, designs that concentrate points on the boundary of the experimental region (such as factorial designs, CCD and the D-optimal design) can yield predictors that perform quite poorly in portions of the experimental region that are observed only sparsely (Santner et al., 2003). On the other hand, space-filling

designs can be a systematic way of discovering surprising behaviours scientifically, because they exercise their code over the entire range of each input variable. The simplest strategy for filling an experimental region E is to select a simple random sample of points from it. However, when the sample size is small and the experimental region is high-dimensional, Latin Hypercube designs are preferable (Santner et al., 2003). Assume that the input variables investigated are independently distributed. In order to derive a Latin Hypercube sample of size S from the experimental region E, the domain of each input variable is divided into S intervals. The set of all possible Cartesian products of these intervals constitutes a partitioning of the p-dimensional sample space into S^p cells, in such a way that the projections of the center of each cell onto each axis are spread uniformly across the axis. The sample is then obtained by selecting S data points such that each row and each column contains only 1 design point (McKay et al., 1979; Santner et al., 2003; Fang et al., 2006).

Along with sampling methods, space-filling designs based on distance measures have also been developed. Their purpose is to obtain designs which have points that are not too close together. One such example is the maximin distance design, which employs the rth order distance between two points $\boldsymbol{w}$ and $\boldsymbol{x}$ in the experimental region E. One important distance measure is defined as $\rho_d(\boldsymbol{w},\boldsymbol{x}) = \left[\sum_{j=1}^{p} \|w_j - x_j\|\right]^{1/r}$, which corresponds to Rectangular distance for $r = 1$, and Euclidean distance for $r = 2$ (Santner et al., 2003). A third type of space-filling design is the uniform design, which consists of a set of points that are distributed uniformly within the experimental region E. Uniform designs are evaluated by comparing the distribution of the design's points with the uniform distribution through a measure of discrepancy (Santner et al., 2003). Typically, the Kolmogorov-Smirnov statistic $\sup_{\boldsymbol{x}\in E} \|F_n(\boldsymbol{x}) - F(\boldsymbol{x})\|$ is used, where F_n is the empirical function of the points in the design D (Santner et al., 2003). Various other space-filling designs have been developed as well, including sphere-packing (Johnson et al., 1990) and lattice (Bates et al., 1996).

3.1.2 Multi-response designs

The design problem is more complex in the presence of multiple responses than in the single response case. As was stated by Khuri and Cornell (1996), a design that is efficient for one response may not be efficient for the remaining responses; therefore, the choice of a design should be based on a criterion which incorporates measures of efficiency that pertain to all of the responses. The results in the area of multi-response designs are restricted to criteria for parameter estimation, design robustness, and a multivariate lack of fit in the presence of linear multi-response models (Khuri and Cornell, 1996). The single-response D-optimal design has been generalized to enable multi-response D-optimal designs to be obtained; however, the computational cost of this procedure limits its applicability (Khuri and Cornell, 1996). Draper and Hunter (1966) proposed a

Bayesian approach to obtaining a design criterion for the selection of s additional experimental runs after S runs have already been performed (Khuri and Cornell, 1996).

3.2 Modeling techniques

Although polynomial models are convenient and generally useful, they are frequently unappealing in NSRSM. Indeed, when a more global representation of the system is desired, the problems of complexity and non-linearity arise (O'Connell and Wolfinger, 1997). Furthermore, the fit of polynomials to the data tends to be erratic near the boundaries, and the necessary extrapolation at the corners of the design can be dangerous (Hastie et al., 2009). As a consequence, first- and second-order linear models give way to highly adaptive models that are better suited to complex non-linear responses and high dimensionality of the input space (Bates et al., 1996). The most popular methods for response modeling in NSRSM are Gaussian stochastic process (GP) models, thin plate splines, and neural networks (Myers et al., 2004). A less diffuse but still relevant approach is the Multivariate Adaptive Regression Splines (MARS) model (Friedman, 1991).

All of the aforementioned modeling techniques move beyond linearity by augmenting or replacing the vector of inputs $\boldsymbol{x}$ with additional variables; such additional variables are transformations of $\boldsymbol{x}$. In this way, a new space of derived input features is obtained and a linear model is subsequently fitted. The great advantage of this approach is that the models obtained are linear in the derived input features (Hastie et al., 2009). The resulting model is a smooth curve obtained through the use of a spline function. The approach, referred to as spline-based smoothing, typically consists of finding the function that, from among all of the functions $f(\boldsymbol{x})$ with two continuous derivatives, minimizes the penalized residual sum of squares

$$\mathrm{RSS}(f,\lambda) = \sum_{n=1}^{N} (y_n - f(\boldsymbol{x}_n))^2 + \lambda \int f''(t)^2 dt, \tag{3}$$

where λ is a fixed smoothing parameter and f'' indicates the second derivative of f. The first term in equation (3) measures the closeness to the data, while the second term penalizes the curvature in the function and establishes a tradeoff between the two. If $\lambda = 0$, f can be any function that interpolates the data, whereas if $\lambda = \infty$, f is the simple least squares line fit, because no second derivative can be tolerated. For $\lambda \in (0, \infty)$, the optimal function varies from very rough to very smooth. The penalty term translates to a penalty on the spline coefficients; this has the effect of shrinking the coefficients toward a linear fit and avoiding over-parametrization (Hastie et al., 2009).

3.2.1 Multivariate Adaptive Regression Splines

Multivariate adaptive regression splines (MARS) is an adaptive procedure for regression which was introduced by Friedman (1991). The model is built using a collection of piecewise-linear basis functions

$$C = \{(X_i - \xi)_+,\ (\xi - X_i)_+\}, \tag{4}$$

where X_i is the ith input variable, $\xi \in \{x_{x1i}, x_{x2i}, \ldots, x_{xNi}\}$ are the associated knots, and $u_+ = \max(u, 0)$ indicates the positive part.

The MARS model is expressed as $f(X) = \beta_0 + \sum_{m=1}^{M} \beta_M h_m(X)$, where β_m are the coefficients, and each $h_m(X)$ is either a function from the collection C or a product of two or more such functions (Hastie et al., 2009). Initially, a large model is built (forward basis selection) by iteratively investigating new terms of the form $\hat{\beta}_{M+1} h_l(X) \cdot (X_j - t)_+ + \hat{\beta}_{M+2} h_l(X) \cdot (t - X_j)_+$, where $\hat{\beta}_{M+1}$ and $\hat{\beta}_{M+2}$ are least squares estimates of the model coefficients, and h_l is one of the transformations belonging to the collection C that are already part of the model. At each step, the term that produces the largest decrease in the training error is added to the model. Once the model contains a preset maximum number of terms, the size of the model is reduced iteratively (backward deletion) in order to avoid overfitting; that is, the term whose removal causes the smallest increase in the residual squared error is deleted from the model. The entire procedure results in an estimated best model f_λ of each size λ (number of terms). The optimal model size λ is estimated by minimizing the generalized cross-validation (GCV) criterion,

$$\text{GCV} = \frac{\sum_{n=1}^{N} (y_n - \hat{f}_\lambda(\boldsymbol{x}_n))^2}{(1 - M(\lambda)/N)^2},$$

where $M(\lambda)$ is the effective number of parameters in the model, namely $M(\lambda) = r + cK$; here, r is the number of linearly independent basis functions in the model, K is the number of knots selected in the forward process, and c is a penalty term (usually, c is either 3 or 2) (Hastie et al., 2009). In the presence of multiple responses, the MARS procedure builds a set of models (one model per output variable) simultaneously. Each model uses the same set of basis functions, but different coefficient values. In order to select the best set of MARS terms across all responses (Milborrow, 2009), the GCV and RSS values are averaged across the responses.

3.2.2 Thin Plate Splines

Consider the response variable Y and the p explanatory variables $X_1, \ldots, X_p$ with realizations $\boldsymbol{y} = (y_1, \ldots, y_N)^T$ and $\boldsymbol{X} = (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_N^T)$. Then, the pairs $(\boldsymbol{x}_n^T, y_n)$, $n = 1, \ldots, N$, are observed with $y_n \in \mathbb{R}$ and $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{np})^T \in \mathbb{R}^p$.

A thin plate spline is a smooth surface of the form

$$f(\boldsymbol{X}) = \beta_0 + \beta^T \boldsymbol{X} + \sum_{n=1}^{N} \alpha_n h_n(\boldsymbol{X}), \tag{5}$$

where $h_n(\boldsymbol{X}) = \eta(\|\boldsymbol{X} - \boldsymbol{x}_n\|)$ and $\eta(z) = z^2 \log z^2$. The model in equation (5) is obtained by solving

$$\min_f \sum_{n=1}^{N} \{y_n - f(\boldsymbol{x}_n)\}^2 + \lambda \cdot J|f|, \tag{6}$$

where $J|f|$ is a penalty function. One example of such a penalty function in the 2-dimensional case is

$$J|f| = \int\int_{\mathbb{R}^2} \left[\left(\frac{\theta^2 f(\boldsymbol{X})}{\theta X_1^2} \right)^2 + \left(\frac{\theta^2 f(\boldsymbol{X})}{\theta X_1 X_2} \right)^2 + \left(\frac{\theta^2 f(\boldsymbol{X})}{\theta X_2^2} \right)^2 \right] dX_1 dX_2, \tag{7}$$

where $\boldsymbol{X} = (X_1, X_2)$ is a 2-dimensional vector of explanatory variables. The parameter λ in equation (5) controls the smoothness of the resulting surface: as $\lambda \to 0$, the solution approaches an interpolating function; as $\lambda \to \infty$, the solution approaches the least squares plane. For intermediate values of λ, the solution can be represented as a linear expansion of basis functions whose coefficients are obtained by a form of generalized ridge regression (Hastie et al., 2009). Radial basis function approximations have been shown to produce good fits for both deterministic and stochastic response functions (Jin et al., 2001).

3.2.3 Gaussian Stochastic Process Model

The use of Gaussian stochastic process models (GP) was introduced in the RSM field in 1989 (Sacks et al., 1989). The GP model can be both deterministic and stochastic. The deterministic GP model is typically employed when computer experiments are carried out; because of the absence of measurement error, the estimated response surface interpolates the observations, meaning that the mean square prediction error at an experimental point $\boldsymbol{x}_n$ is zero, namely $\hat{y}(\boldsymbol{x}_n) = y(\boldsymbol{x}_n)$ (Sacks et al., 1989).

The deterministic model is given by the sum of two terms

$$\boldsymbol{y}(\boldsymbol{X}) = \sum_{i=1}^{p} \beta_i f_p(\boldsymbol{X}) + Z(\boldsymbol{X}). \tag{8}$$

The first term $\sum_{i=1}^{p} \beta_i f_p(\boldsymbol{X})$ is assumed to be a linear model, while the second term $Z(\boldsymbol{X})$ is a systematic departure from the assumed linear model and is treated as a realization of a stochastic process (random function) in which the covariance structure of Z is related to the smoothness of the response (Sacks et al., 1989).

The covariance structure of the stochastic process Z is given by $\text{cov}(Z(\boldsymbol{t}), Z(\boldsymbol{u})) = \sigma_Z^2(R(\boldsymbol{t}, \boldsymbol{u}))$, where $\boldsymbol{t}$ and $\boldsymbol{u}$ are p-dimensional inputs $\boldsymbol{t} = (t_1, \ldots, t_p)$ and $\boldsymbol{u} = (u_1, \ldots, u_p)$, σ_Z^2 is a scale factor, and R is the correlation function. A covariance function with some derivatives is used in the case of a smooth response, whereas a covariance function with no derivatives is appropriate in the presence of an irregular response (Sacks et al., 1989). Usually, the correlation function is considered to be the product of one-dimensionally stationary correlations, namely $R(\boldsymbol{t}, \boldsymbol{u}) = \prod R_i(t_i - u_i)$.

Of special interest are correlations of the form

$$R(\boldsymbol{t}, \boldsymbol{u}) = \prod \exp(-\theta |t_i - u_i|^d), \tag{9}$$

where $0 < d \leq 2$, for which the parameter θ defines the correlation structure of Z. In particular, large values of θ result in small correlations between observations, whereas small values of θ results in large correlations. Therefore, the parameter plays a critical role in prediction (Sacks et al., 1989).

The predictor $\hat{\boldsymbol{y}}(\boldsymbol{X})$ is usually the best linear unbiased predictor (BLUP) of $\boldsymbol{y}(\boldsymbol{X})$, and is obtained by minimizing its mean squared prediction error (MSPE). Either cross validation or maximum likelihood estimation (MLE) can be used to estimate the parameters (θ and p) if a sufficiently large set of data is available.

In the case of maximum likelihood estimation, $E[Z(\boldsymbol{X})]$ is typically assumed to be zero and the joint distribution of the $Z(\boldsymbol{X})$ is assumed to be Gaussian (i.e., Z is a Gaussian process). The likelihood is a function of the βs, the process variance σ_Z^2, and the correlation parameters. Given the correlation parameters, the MLE of the βs and σ_Z^2 are obtained, and the problem is reduced to minimizing $(\det R)^{1/n} \hat{\sigma}_Z^2$, which is a function of only the correlation parameters and the data (Sacks et al., 1989).

Once the parameters have been specified, equation (9) can be used in equation (8) to provide predictions of $y(\boldsymbol{x})$ at an untried experimental point $\boldsymbol{x}$ (Sacks et al., 1989). The parameter θ is generally not available prior to data collection, and a design strategy that will perform well for a wide range of unknown values of θ would be advisable. Therefore, it is recommended that a robustness study be performed at the design stage, in order to select a robust value of θ, say θ_A. This robust θ_A can also be used if the data collected are inadequate for estimating θ (Sacks et al., 1989).

The stochastic GP model is obtained by adding the random error term $\varepsilon(\boldsymbol{X})$ to the deterministic model in equation (8), such that

$$\boldsymbol{y}(\boldsymbol{X}) = \sum_{i=1}^{p} \beta_i f_p(\boldsymbol{X}) + Z(\boldsymbol{X}) + \varepsilon(\boldsymbol{X}). \tag{10}$$

Random errors which are independent of both each other and Z are incorporated by employing a covariance structure reflecting $\text{var}(Y(t)) = \sigma_Z^2 + \sigma_\varepsilon^2$ (Sacks et al., 1989). The covariance structure is given by $\text{cov}(Y_i, Y_j) = \frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_\varepsilon^2} \cdot R(\boldsymbol{x}_i, \boldsymbol{x}_j)$ (Huang et al., 2006).

3.2.4 Neural Networks

Neural networks are powerful learning methods that are inspired by the way in which the human brain processes information (Hastie et al., 2009). Neural networks consist of a large number of highly interconnected processing elements (neurons). Neural networks undergo training and learning processes during which the synaptic connections between the neurons are adjusted (Balkin and Lin, 2000).

Mathematically, neural networks are nonlinear statistical models that are obtained in two steps (Hastie et al., 2009). First, linear combinations of the inputs are extracted as derived features; second, the target is modeled as a nonlinear function of these features (Hastie et al., 2009).

Let $\boldsymbol{\alpha}_m$, $m = 1,\ldots,M$, be p-dimensional vectors of unknown parameters, and let Z_m be the features derived. The most widely used neural network is the single hidden layer back-propagation network (known as the single layer perceptron), which consists of a single layer of hidden units. However, neural networks with more hidden layers also exist. The hidden units are in the middle of the network and serve to compute the derived features

$$Z_m = \sigma(\alpha_{0m} + \boldsymbol{\alpha}_m^T \boldsymbol{X}), \tag{11}$$

with $m = 1,\ldots,M$. The term $\sigma(\nu)$ in equation (11) is referred to as the activation function; usually, the sigmoid $\sigma(\nu) = 1/(1 + e^{-\nu})$ is used as a smooth threshold function.

Assume that q response variables $Y_1,\ldots,Y_q$ are investigated. Each output Y_j, $j = 1,\ldots,q$, is obtained as a function of linear combinations of $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_M)$ and $Y_j = f_j(\boldsymbol{X}) = g_j(\beta_{0j} + \beta_j^T \boldsymbol{Z})$, with $j = 1,\ldots,q$. Therefore, the final response is given by $Y_j = f_j(\boldsymbol{X}) = g_j(\beta_{0j} + \beta_j^T \sigma(\alpha_{0m} + \boldsymbol{\alpha}_m^T \boldsymbol{X}))$.

The parameters of a neural network are known as weights. They consist of $M(p+1)$ weights $\{\alpha_{0m}, \boldsymbol{\alpha}_m\}$, with $m = 1,\ldots,M$, and $q(M+1)$ weights $\{\beta_{0j}, \beta_j\}$, with $j = 1,\ldots,q$. Let θ denote the complete set of weights. Given a set of training data $(\boldsymbol{x}_n^T, \boldsymbol{y}_n^T)$, $n = 1,\ldots,N$, where $\boldsymbol{x}_n = (x_{n1},\ldots,x_{np})^T$ and $\boldsymbol{y}_n = (y_{n1},\ldots,y_{nq})^T$, with $n = 1,\ldots,N$, the weights are estimated by minimizing the penalized error function

$$R(\theta) = \sum_{j=1}^{q} \sum_{n=1}^{N} (y_{nj} - f_j(\boldsymbol{x}_n))^2 + \lambda \cdot J(\theta), \tag{12}$$

where $\lambda \geq 0$ is a tuning parameter and $J(\theta)$ is a penalty term used to avoid overfitting. The tuning parameter λ is typically estimated by cross-validation and its value controls the roughness of the

fitted function; in particular, large values of λ tend to shrink the weights toward zero. The penalty term $J(\theta)$, referred to as weight decay, is the sum of squares of the weights θ (Hastie et al., 2009).

3.3 Optimization procedures

3.3.1 Single-response optimization

Model-based optimization methods make use of model fitting to identify the system optimum. The most common approaches are "one-shot" methods that fit the system response with a regression model and determine the optimal solution from the optimization of the model itself (Huang et al., 2006). The aim of "one-shot" approaches is to provide a model that approximates the unknown input-output relationship closely over the entire domain of feasibility. Large numbers of observations are usually required to achieve this, and the methodology becomes inefficient or even unaffordable when applied to physical experimentation. Sequential procedures have therefore been developed to address this challenge (Sacks et al., 1989). Sequential procedures rely on smaller numbers of evaluated data points by relaxing the need to get a good fit throughout the entire operability region and aiming for good predictions in the neighborhood of the optimum (Sacks et al., 1989). Both local and global sequential optimization procedures have been developed; our focus here is on global methodologies.

Existing sequential global optimization techniques have been developed mainly in the deterministic simulation framework, and the research literature has not considered their adaptation to the stochastic context (Huang et al., 2006). Overall, sequential global optimization methods can be classified into two broad classes, namely two-stage and one-stage approaches (Jones, 2001). Two-stage approaches consist of first fitting a model to the observed data and estimating all of the relevant parameters (stage one); then using the fitted model to determine additional design points that have not been investigated yet (stage two). One-stage approaches, on the other hand, aim to estimate the model parameters and, at the same time, determine new design points to investigate (Jones, 2001). In what follows, the main two-stage and one-stage sequential optimization techniques are reviewed; the following discussion is based mainly on the work by Jones (2001).

The simplest two-stage approach consists of taking the current optimal point as an additional design point to be investigated. This can result in a premature convergence to a local maximum (or minimum), because the investigation is limited to the area surrounding the current optimum. With global optimization, it is crucial to achieve a proper balance between the exploitation of the current optimum area and the exploration of the remaining part of the search space (Tripathi et al., 2007). This can be achieved by forcing the optimization procedure to also pay attention to parts of the space that are relatively unexplored. Three such procedures make use of the de-

terministic Gaussian stochastic process model; indeed, the GP model provides a measure of the model uncertainty that can be optimized in order to direct the search towards unsampled regions as well.

The first of these approaches aims to minimize a statistical lower bound of the form $\hat{y} - ks(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^p$ is a design point in the experimental region E, s is the root mean squared error or standard error, and k is a constant value. The search in high standard error regions is emphasized when $k > 0$ is used (Cox and John, 1997).

The second two-stage approach consists of finding the design point that maximizes the probability of improving the approximating function beyond some target T. Let f_{min} be the current best function value. If the aim is to minimize the response function, then the target T will be some number $T < f_{min}$. Assume for example that a 25% improvement is desired; then, the target is set to $T = f_{min} - 0.25 \cdot |f_{min}|$.

At any given point, the value of the approximating function is considered to be the realization of a random variable Y with mean $\hat{y}(\boldsymbol{x})$ and standard error $s(\boldsymbol{x})$. The probability of improvement is the probability that $Y(\boldsymbol{x}) \leq T$. If Y is assumed to be normally distributed, the probability of improvement is given by $\text{PI} = \Phi\left(\frac{T - \hat{y}(\boldsymbol{x})}{s(\boldsymbol{x})}\right)$, where $\Phi(\cdot)$ is the Normal cumulative distribution function. As the function is sampled more and more around the current best point, the probability of improvement will be small, and the standard error will become smaller and smaller. Eventually, the probability of an improvement around the current best point will become so small that the search will be driven to where the standard error is higher.

This approach is extremely sensitive to the choice of the target T. Indeed, for an excessively small desired improvement, a nearly exhaustive search around the current best point is required before a more global search in begun. On the other hand, with an excessively high desired improvement, the search will be excessively global and the algorithm will be slow to fine-tune any promising solutions.

The use of several target values has therefore been proposed in order to overcome the sensitivity of the approach to the choice of T. In particular, low, medium, and high levels of desired improvement are specified. This results in the selection of several search points, allowing for both local and global searches. In particular, the global search identifies the basin of convergence of the global minimum, then the local search begins to fine-tune the solution.

The third two-stage approach consists of computing the expected improvement when sampling at a given point. Such approach was popularised by the introduction of the Efficient Global Optimization (EGO) algorithm (Jones et al., 1998). Similarly to the previous approach, $Y(\boldsymbol{x})$ is a random variable describing the uncertainty about the value of the function at a point $\boldsymbol{x}$. $Y(\boldsymbol{x})$ is

typically assumed to be normally distributed, with its mean and variance given by the GP predictor (i.e., $\hat{y}(\boldsymbol{x})$ and $s^2(\boldsymbol{x})$). Let f_{min} be the current best function value and I be the desired improvement; then, the improvement I will be achieved if $Y(\boldsymbol{x}) = f_{min} - I$. The likelihood of achieving this improvement is given by

$$L(f_{min}, s; y) = \frac{1}{\sqrt{(2\pi s(\boldsymbol{x}))}} \exp\left[\frac{(f_{min} - I - \hat{y}(\boldsymbol{x}))^2}{2s^2(\boldsymbol{x})}\right], \tag{13}$$

whereas the expected improvement is simply the expected value of the improvement in equation (13), namely

$$E(I) = \int_0^\infty I \left\{ \frac{1}{\sqrt{(2\pi s(\boldsymbol{x}))}} \exp\left[\frac{(f_{min} - I - \hat{y}(\boldsymbol{x}))^2}{2s^2(\boldsymbol{x})}\right] \right\} dI. \tag{14}$$

Unlike the probability of improvement, the expected improvement does not require the specification of the desired improvement, and therefore it overcomes the sensitivity issue related to the choice of the target T. Furthermore, the procedure stops automatically when the expected improvement is less than a predefined small positive number. Similarly to the probability of improvement, the expected improvement might take a long time to find the global minimum. This is because an exhaustive local search around the initial best point is required before the algorithm moves to a more global search. Indeed, the estimates of the standard error which are obtained initially are very small, and therefore, only points that are close to the current best point have a high expected improvement. In the presence of observation noise, both the model and optimization criteria must be adapted in order to address stochastic systems. Usually, the GP model in equation (10) is employed; the model parameters are derived from the noisy data, whereas the observation noise variance σ^2 is treated as an additional parameter to be optimized. For the optimization criterion, the Augmented Expected Improvement (Huang et al., 2006), the multi-fidelity optimization (Forrester et al., 2007) and the quantile-based Expected Improvement (Picheny et al., 2013) have been developed.

When the initial sample is sparse, two-stage methods can be deceived. Indeed, stage two relies on the model obtained in stage one. If the model chosen in stage one provides a highly misleading view of the true optimization function, the procedure may fail to converge to the true system optimum, as the algorithm either stops prematurely or becomes excessively local in selecting the design points. One-stage methods (Gutmann, 2001) escape this pitfall by estimating the model parameters based on information beyond the observed sample. Indeed, it is assumed that the model goes through the point $(\boldsymbol{x}^*, y^*)$, where y^* is a desirable value for the objective function (e.g., a benchmark to competitive products) and $\boldsymbol{x}^*$ is the design point at which it is hypothesized that the target y^* is achieved. To evaluate the stated hypothesis, a conditional likelihood is computed, namely the likelihood of the observed data, conditional on the assumption that the surface goes through the point $(\boldsymbol{x}^*, y^*)$. If a deterministic GP model is used, the conditional likelihood is given

by

$$L(m, \boldsymbol{C}; y) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}|C|^{1/2}} \exp\left[\frac{-(y-m)^T \boldsymbol{C}^{-1}(y-m)}{\sigma^2}\right], \tag{15}$$

where m and $\boldsymbol{C}$ are the conditional mean and the correlation matrix. Thus, the model parameters are estimated by maximizing the conditional likelihood in equation (15). A measure of the credibility of the stated hypothesis is provided as well. It is worth noting that the computational cost of the procedure does increase significantly in the case of an unknown or unclear target, because the process must be repeated using several search points. One-stage methods have been combined with both the Gaussian process model and the thin plate splines model (Gutmann, 2001). The second class of models provides significant improvements from a computational perspective.

3.3.2 Multi-response optimization

A move from single-response to multi-response systems increases the complexity of the optimization problem (Khuri and Cornell, 1996). Indeed, the conditions that are optimal for one response may be far from optimal or even physically impractical for other system responses. It is actually rare for all of the response variables to achieve their respective optima under exactly the same conditions (Khuri and Cornell, 1996).

There are four main ways of dealing with multiple responses in RSM. The simplest strategy consists of building an appropriate response surface model for each response, overlaying the obtained contour plots, and locating a region where the conditions are 'near' optimal for all of the responses (Montgomery, 2009; Khuri and Cornell, 1996). However, this procedure only suits situations where limited numbers of both inputs and response variables are investigated. A different approach involves the optimization of a dual response system which consists of a primary response and a secondary response (Myers and Carter, 1973). The methodology aims to find the input conditions that optimize the primary response, subject to the condition that the secondary response takes on some specified or desirable values. This results in a constrained optimization problem, the solution of which is similar to that of a ridge analysis (Khuri and Cornell, 1996).

A similar method which has its roots in the multi-objective optimization community is the lexicographic approach. It consists of assigning different priorities to different objectives, after which the objectives are optimized according to the predefined priority order. In particular, the performance measures of two or more candidate design points are compared in terms of the highest-priority objective. If one design point is significantly better, it is chosen; otherwise, the performance measures of the candidate design points are compared with respect to the second objective. This process is then repeated until either a winner design point is found or all of the criteria have been used (Freitas, 2004).

The third methodology used in RSM is the desirability function approach (Harrington, 1965). Similarly to the previous methodologies, a separate model is initially built for each investigated response, then the predicted response function is transformed into a desirability function h_j, $0 \leq h_j \leq 1$, which increases as the desirability of the corresponding property increases (Khuri and Cornell, 1996). Various different desirability functions have been developed, allowing a high degree of flexibility to be achievable. One such example is the exponential-type transformations $h_j = \exp(-|\hat{Y}_j|^s)$, where s is a user-specified positive number; the following functions are also extremely common:

$$h = \begin{cases} \left(\frac{\hat{Y}-A}{B-A}\right)^s, & A \leq \hat{Y} \leq B \\ \left(\frac{\hat{Y}-C}{B-C}\right)^s, & B \leq \hat{Y} \leq C \end{cases}, \tag{16}$$

where h assumes the value 0 if $\hat{Y} > C$ or $\hat{Y} < A$ (Khuri and Cornell, 1996; Myers and Montgomery, 1995). Finally, the individual desirability functions are combined using the geometric mean $H = \left\{h_1 \cdot h_2 \dots h_q\right\}^{1/q}$; thus, a single function assessing the overall desirability is obtained (Khuri and Cornell, 1996), and the design point that maximizes H is chosen (Myers and Montgomery, 1995). Because the overall desirability is a well-behaved continuous function of the input variables, the optimization process is relatively simple and is performed using univariate optimization techniques (Khuri and Cornell, 1996). Ideally, the value of H should be close to 1, because this indicates that all of the responses are in the desirable range simultaneously (Myers and Montgomery, 1995).

The fourth methodology is the generalized distance approach (Khuri and Conlon, 1981). The basic assumptions are that all of the responses depend on the same set of input variables, and can be represented by polynomial regression models of the same degree within the experimental region. Each individual polynomial model is optimized separately, and its optimum value φ_j identified. In the presence of q responses, the vector $\hat{\boldsymbol{\phi}} = (\phi_1, \dots, \phi_q)^T$ indicates an "ideal optimum". Such an ideal optimum rarely exists, because the individual optima are seldom attained under the same set $\boldsymbol{x}$ of operating conditions. Therefore, a near optimum is searched for, and conditions on the input variables that are somewhat favourable to all of the responses are explored. The deviation of the compromising conditions from the ideal optimum is measured by means of a distance function between the predicted responses $\hat{\boldsymbol{y}}(\boldsymbol{x})$ and the vector of individual optima $\hat{\boldsymbol{\phi}}$. Usually, the weighted distance $\rho(\hat{\boldsymbol{y}}(\boldsymbol{x}) - \hat{\boldsymbol{\phi}}) = [(\hat{\boldsymbol{y}}(\boldsymbol{x}) - \hat{\boldsymbol{\phi}})^T \mathrm{Var}(\hat{\boldsymbol{y}}(\boldsymbol{x}))^{-1} (\hat{\boldsymbol{y}}(\boldsymbol{x}) - \hat{\boldsymbol{\phi}})]^{1/2}$ is employed (Khuri and Cornell, 1996). An extremely similar approach has also been developed in the multi-objective optimization context, known as the method of distance functions, which consists of combining the multiple objectives $f_j(\boldsymbol{x})$ into a single objective function using

$$W = \left[\sum_{j=1}^{q} |f_j(\boldsymbol{x}) - y_j^*|^r\right]^{1/r}. \tag{17}$$

Here, $1 \leq r < \infty$, $\boldsymbol{x}$ is a point in the experimental region E, and $\boldsymbol{y}^* = (y_1^*, \ldots, y_q^*)$ is a user-defined vector of targets. For $r = 2$, the Euclidean distance is obtained (Srinivas and Deb, 1994).

Two additional methodologies that are commonly employed in the multi-objective optimization field are the weighted-formula approach and the Pareto approach. The weighted-formula approach consists of normalizing the q objective functions, $f_j'(\boldsymbol{x})$, assigning them numerical weights w_j such that $\sum_{j=1}^{q} w_j = 1$, and combining the weighted responses through addition

$$\min f = w_1 f_1'(\boldsymbol{x}) + w_2 f_2'(\boldsymbol{x}) + \cdots + w_q f_q'(\boldsymbol{x}) \tag{18}$$

or multiplication

$$\min f = f_1'(\boldsymbol{x})^{w_1} + f_2'(\boldsymbol{x})^{w_2} + \cdots + f_q'(\boldsymbol{x})^{w_q} \tag{19}$$

(Freitas, 2004).

Unlike the methods which have been described thus far, the Pareto approach does not transform a multi-objective problem into a single-objective problem; instead, it solves the original multi-objective problem directly. This approach is based on the concept of Pareto dominance. If all of the objective functions f_j are for minimization, a solution $\boldsymbol{x}_1$ is said to dominate another solution $\boldsymbol{x}_2$ if and only if $\boldsymbol{x}_1$ is strictly better than $\boldsymbol{x}_2$ with respect to at least one of the objective functions being optimized, and $\boldsymbol{x}_1$ is not worse than $\boldsymbol{x}_2$ with respect to all the objective functions (Freitas, 2004). Mathematically, $\boldsymbol{x}_1$ dominates $\boldsymbol{x}_2$ if and only if $f_j(\boldsymbol{x}_1) \leq f_j(\boldsymbol{x}_2)$ for $j = 1, \ldots, q$ and $f_j(\boldsymbol{x}_1) < f_j(\boldsymbol{x}_2)$ for at least one objective function f_j (Konak et al., 2006).

The Pareto-based approach provides a set of non-dominated solutions that is referred to as the Pareto optimal set. The corresponding set of objective function values is called the Pareto front (Konak et al., 2006). The Pareto optimal set identifies solutions which represent various compromises between the multiple responses; indeed, for these solutions, no objective can be improved without detracting from at least one other objective (Mansouri, 2005). The choice of the best solution is performed a posteriori by selecting it from the Pareto optimal set (Konak et al., 2006).

Usually, Pareto-based optimization makes use of multi-objective algorithms. Heuristic methods are particularly suitable because of their ability to find good (though not necessarily optimal) solutions to difficult combinatorial optimization problems (Myers et al., 2004; Feo and Resende, 1995). Relevant techniques include simulated annealing (Kirkpatrick, 1984), the tabu search (Glover, 1989, 1990) and genetic algorithms (Goldberg, 1989).

The mechanisms underlying the simulated annealing and tabu search algorithms involve refining a single candidate solution. In contrast, genetic algorithms evolve several different candidate solutions (a population) simultaneously (Paterlini and Krink, 2006). This allows for a parallel search that is particularly suitable for both global and multi-objective optimizations (Fonseca

and Fleming, 2005). Genetic algorithms are a stochastic search heuristic which was inspired by Darwinian evolution and genetics (Holland, 1975).

The main idea is to create a population of candidate solutions that is refined iteratively by altering and selecting good solutions. The solutions are altered through both mutation and crossover, which allow for the random exploration of the solutions' local neighbourhoods and the recombination of different solutions' information respectively (Paterlini and Krink, 2006). The goodness of a solution is expressed in terms of one or more fitness functions, which are then used to select good solutions. Often, the fitness functions correspond to the system's response functions. Because such functions are usually either unknown or computationally expensive, surrogate models can be used to approximate them. Surrogate models are constructed from a limited number of data points obtained during one or more generations of a classical evolutionary search. The surrogate model is updated as new data points become available (Zhou et al., 2007). For a comprehensive survey of this topic, see Jin (2005, 2011).

Polynomial models, neural networks, radial basis functions (RBF), and Gaussian process (GP) models are among the most prominent and common techniques used to approximate the fitness function in evolutionary computation. Thus, the connection with (nonparametric and semiparametric) RSM becomes evident. Particle Swarm Optimization and Ant Colony Optimization are other heuristics from the class of population-based procedures. Particle swarm and ant colony optimization are swarm intelligence search heuristics (Jin, 2005). Particle swarm optimization mimics the social behavior of a flock of birds in order to guide swarms of particles towards the most promising regions of the search space (Kennedy and Eberhart, 1995; Jin, 2005). Ant colony optimization is inspired by the foraging behavior of ant colonies (Dorigo, 1992); each ant evaluates its path (solution) according to a fitness function and moves through the search space based on the pheromone trial, a form of indirect communication mediated by a modification of the environment (Dorigo, 1992).

Heuristics require a large number (thousands) of function evaluations. Despite the fact that they are feasible in simulated studies, the application of these techniques to real-world studies is usually impracticable. To overcome this drawback, methodologies such as ParEGO (Knowles and Hughes, 2005) and its variations (Ponweiser et al., 2008; Zhang et al., 2010), and the Evolutionary Model-based Multiresponse Approach (EMMA) have been developed (Villanova et al., 2010; Carta et al., 2011). ParEGO was introduced by Knowles (Knowles and Hughes, 2005); it is the multiresponse version of EGO (Jones et al., 1998). The extension of EGO to the multi-objective case is achieved by converting the q response functions into a single response via a parameterized scalarizing weight vector (Knowles and Hughes, 2005). EMMA aims to find compromise solutions that are as close to a desirable target as possible. It achieves this by combining a Multivariate Adaptive Regression Splines model with a time variant Particle Swarm Optimization heuristic and the gen-

eralized distance approach. The Multivariate Adaptive Regression Splines model approximates the system responses and is used to identify the desirable target. The particle swarm optimization heuristic guides the search for promising solutions by selecting new design points on the basis of the distance functions in equation (17). To avoid the selection of false optima, the response values of the newly identified design points are measured on the true system directly.

4 Statistical challenges and future research

This paper summarizes various statistical design, modeling and optimization techniques that are applicable to the efficient identification of optimal conditions in applications. The review is motivated by the significant methodological developments that have been occurring in many different research fields simultaneously over recent years. Despite their power and increasing versatility, such methodologies are not being exploited adequately in real-world studies.

The main statistical challenges for future research arise from the following issues:

Complexity and High Dimensionality. Contemporary experimentation is dealing increasingly with problems that relate to complex systems; these involve very large numbers of features and variables that can significantly affect the responses to the experimental questions. This high dimensionality is currently one of the most crucial problems in experimentation, modelling and optimisation (Johnstone and Titterington, 2009; Shan and Wang, 2009). The number of experimental points required by classical procedures increases exponentially with the number of input variables (Wang and Shan, 2007), with a subsequent increase in the cost of experimentation and computational resources required. Finding good modeling and optimisation techniques is also very difficult when dealing with high dimensional systems, and often the practice is to reduce the dimensions of the problem *a priori* using *ad hoc* procedures that can be both inaccurate and misleading for the problem (Fan and Lv, 2010; Science, 2011). New directions to head in this area of research involve the construction of integrated approaches, which develop and combine novel procedures according to the sequence of design, modeling and optimisation.

Designs generally involve small sets of data (few experimental runs) relative to the large experimental space. Initial designs are frequently selected under a random sampling procedure, since randomness may allow the exploration of the search space in areas which are not anticipated by prior knowledge but where relevant information may reside. From the initial (explorative) designs, a sequence of more and more informative designs are then derived under different heuristic procedures, such as evolutionary algorithms and stochastic local search algorithms (Blum et al., 2011; Baragona, 2011). Different sampling procedures and

adaptive processes for the variation of the dataset require further investigation to improve the search procedures employed during the optimization. The recent contributions on supersaturated designs may represent an interesting way to tackle the problem in this context as well (Claeys-Bruno et al., 2011; Georgiou, 2012).

Modelling high dimensional data with few observations is a growing field of research and a very challenging problem. Abandoning the assumptions of classical statistical modeling, where the number of observations is supposed to be large and the number of variables small, also involves abandoning the carefully accumulated theoretical results on bias, accuracy and efficiency, and most of the asymptotic results. New directions of research are based mainly on the sparse nature of the problems, and consider the hypothesis that a small set of variables can affect the experimental response (Meinshausen and Bühlmann, 2010). From the Lasso (Least Absolute Shrinkage and Selection Operator) family of techniques, which involve penalized maximum likelihood estimators (Griffin and Brown, 2012, 2011; Mai and Zou, 2012; Bien and Tibshirani, 2011), other approaches are now emerging based on composite probability prior distributions ("slab and spike variable selection") (Mai and Zou, 2012), and on structured Bayesian hierarchical models (Griffin and Brown, 2012, 2011). The ways in which models can affect the design, providing extra information, forms an open and very interesting area. From a different perspective, another approach which is growing in both research and interest is the decomposition methodology (Li et al., 2009), which involves the reformulation of the original problem into a set of independent sub-problems of lower dimensionality.

Optimization, both single and multi-objective optimization, is taking shape in the context of heuristic approaches such as "model-based optimization". In fact, the search for the optimum value in high dimensional settings is increasingly being derived using both the rules of the relevant heuristic search (evolution, swarm intelligence, ant colony, or others) and the information from the statistical model being estimated on the experimental responses. The problem of high dimensionality is further accentuated when the Pareto approach is adopted in the presence of multiple responses. Researchers are increasingly developing multi-objective optimization procedures for dealing with large-scale problems (Ta et al., 2011; Guo et al., 2008; Kandil et al., 2010; Santos et al., 2010); however, further research is required to cope with problems which feature a limited budget of available resources.

Pareto-based procedures. Far too often, multi-objective problems are transformed into problems with single objectives by combining the multiple responses of interest into a single objective function. If the multi-objective nature of the problem is to be preserved, it is necessary to use either multiple models or, preferably, a single multivariate model. Indeed, multivariate modeling accounts not only for the input-output relationship, but also for the correlation

between the responses (Frank and Friedman, 1993). Similarly, multi-objective optimization criteria are required to drive the selection of new experimental points, while taking into consideration the multidimensionality of the response space. Research in this direction was suggested by Couckuyt et al. (2012).

Robust Parameter Design. Robust Parameter Design is a field of DOE dealing with the simultaneous optimization of both the product characteristics and the process variability. The Robust Parameter Design strategy was originally proposed by Taguchi (1991); subsequently, the controversy about Taguchi's methodology boosted the development of a new series of statistically sound alternative approaches (Myers et al., 1992). The field of research of Robust Parameter Design has received a great deal of attention in recent years as well (Zhou et al., 2013; Sharda and Banerjee, 2013; Ren et al., 2013; Sibalija and Majstorovic, 2012; Tang et al., 2012; Nakagawa and Kirikoshi, 2012; Holimchayachoikul et al., 2011). The focus on multi-objective optimization is significant, and relevant attention is devoted to the use of the Pareto approach. Robust Parameter Design procedures have benefited greatly from developments in Pareto-based methodologies.

Benchmarking and free software capability. Ideally, benchmarks should reproduce the features of the real-world problem at hand closely. Customizable problem dimensionalities and noise levels for both the input and output variables are examples of desirable features. Despite the fact that advances in such directions have been achieved (Huband et al., 2005), additional obstacles to the straightforward comparison of the performances of model-based optimization procedures remain. One such obstacle is the lack of both test problems and optimization algorithms in the form of freely available packages or source code. The research community would surely benefit from investments in existing free computing platforms (R Development Core Team, 2009).

References

Balkin, S. D. and D. K. J. Lin (2000) A neural network approach to response surface methodology, *Communications in Statistics – Theory and Methods*, **29**(9 & 10), 2215–2227.

Baragona, R. (2011) *Evolutionary statistical procedures*, Springer-Verlag: Berlin, Heidelberg.

Bartz-Beielstein, T., C. W. G. Lasarczyk and M. Preuss (2005) Sequential parameter optimization, in *2005 IEEE Congress on Evolutionary Computation*, vol. 1, pp. 773–780, IEEE, Edinburgh, Scotland, UK.

Bates, R. A., R. J. Buck, E. Riccomagno and H. P. Wynn (1996) Experimental design and observation for large systems (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**(1), 77–94.

Bien, J. and R. J. Tibshirani (2011) Sparse estimation of a covariance matrix, *Biometrika*, **98**(4), 807–820.

Blum, C., J. Puchinger, G. R. Raidl and A. Roli (2011) Hybrid metaheuristics in combinatorial optimization: A survey, *Applied Soft Computing*, **11**(6), 4135–4151.

Box, G. E. P. and K. B. Wilson (1951) On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society, Series B (Methodological)*, **13**(1), 1–15.

Büche, D., N. N. Schraudolph and P. Koumoutsakos (2004) Accelerating evolutionary algorithms with Gaussian process fitness function models, *IEEE Transactions on Systems, Man and Cybernetics*, **35**, 183–194.

Carta, D., L. Villanova, S. Costacurta, A. Patelli, I. Poli, S. Vezzù, P. Scopece, F. Lisi, K. Smith-Miles, R. J. Hyndman, A. J. Hill and P. Falcaro (2011) Method for optimizing coating properties based on an evolutionary algorithm approach, *Analytical chemistry*, **83**(16), 6373–80.

Chen, V. C., K.-L. Tsui, R. R. Barton and M. Meckesheimer (2006) A review on design, modeling and applications of computer experiments, *IIE Transactions*, **38**(4), 273–291.

Claeys-Bruno, M., M. Dobrijevic, R. Cela, R. Phan-Tan-Luu and M. Sergent (2011) Supersaturated designs for computer experiments: Comparison of construction methods and new methods of treatment adapted to the high dimensional problem, *Chemometrics and Intelligent Laboratory Systems*, **105**(2), 137.

Couckuyt, I., D. Deschrijver and T. Dhaene (2012) Towards Efficient Multiobjective Optimization: Multiobjective statistical criterions, in *2012 IEEE Congress on Evolutionary Computation*, pp. 1910–1917, Ieee.

Cox, D. D. and S. John (1997) Multidisciplinary design optimization: State of the art, in N. Alexandrov and M. Hussaini (eds.) *SDO: A statistical method for global optimization*, pp. 315–329, SIAM, Philadelphia.

Dellino, G., J. P. Kleijnen and C. Meloni (2010) Robust optimization in simulation: Taguchi and Response Surface Methodology, *International Journal of Production Economics*, **125**(1), 52–59.

Dorigo, M. (1992) *Optimization, learning and natural algorithms (in Italian)*, Master's thesis, Dipartimento di Elettronica, Politecnico di Milano, Milan, Italy.

Draper, N. R. and W. G. Hunter (1966) Design of experiments for parameter estimation in multiresponse situations, *Biometrika*, **53**(3 & 4), 525–533.

Fan, J. and J. Lv (2010) A selective overview of variable selection in high dimensional feature space, *Statistica Sinica*, **20**, 101–148.

Fang, K.-T., R. Li and A. Sudjianto (2006) *Design and modeling for computer experiments*, Chapman & Hall, Boca Raton, FL.

Feo, T. A. and M. G. C. Resende (1995) Greedy randomized adaptive search procedures, *Journal of Global Optimization*, **6**, 109–133.

Fonseca, C. and P. J. Fleming (2005) An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Computation*, **3**(1), 1–16.

Forrester, A. I., A. Sóbester and A. J. Keane (2007) Multi-fidelity optimization via surrogate modelling, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **463**(2088), 3251–3269.

Frank, I. E. and J. H. Friedman (1993) A statistical view of some chemometrics regression tools (with discussion), *Technometrics*, **35**(2), 109–148.

Freitas, A. A. (2004) A critical review of multi-objective optimization in data mining: a position paper, *ACM SIGKDD Explorations Newsletter*, **6**(2), 77–86.

Friedman, J. H. (1991) Multivariate adaptive regression splines (with discussion), *The Annals of Statistics*, **19**(1), 1–141.

Georgiou, S. D. (2012) Supersaturated designs: A review of their construction and analysis, *Journal of Statistical Planning and Inference*, **142**(10).

Glover, E. (1989) Tabu search—part I, *ORSA Journalon Computing*, **1**, 190–206.

Glover, E. (1990) Tabu search—part II, *ORSA Journalon Computing*, **2**, 4–32.

Goldberg, D. E. (1989) *Genetic algorithms in search, optimization and machine learning*, Addison-Wiley.

Grierson, D. and W. Pak (1993) Optimal sizing, geometrical and topological design using a genetic algorithm, *Structural Optimization*, **6**(3), 151–159.

Griffin, J. E. and P. J. Brown (2011) Bayesian Hyper-Lassos With Non-Convex Penalization, *Australian & New Zealand Journal of Statistics*, **53**(4), 423–442.

Griffin, J. E. and P. J. Brown (2012) Structuring shrinkage: some correlated priors for regression, *Biometrika*, **99**(2), 481–487.

Guo, J., Y. Wang, S. Member, K.-s. Tang, S. Chan, E. W. M. Wong, S. Member, P. Taylor and M. Zukerman (2008) Evolutionary Optimization of File Assignment for a Large-Scale Video-on-Demand System, *IEEE Transactions on Knowledge and Data Engineering*, **20**(6), 836–850.

Gutmann, H. (2001) A radial basis function method for global optimization, *Journal of Global Optimization*, **19**(3), 201–227.

Harrington, E. C. (1965) The desirability function, *Industrial Quality Control*, **21**, 494–498.

Hastie, T., R. Tibshirani and J. Friedman (2009) *The elements of statistical learning*, Springer, New York, 2nd ed.

Hill, W. J. and W. G. Hunter (1966) A Review of Response Surface Methodology: A Literature Survey, *Technometrics*, **8**(4), 571–590.

Holimchayachoikul, P., K. Leksakul and G. Guizzi (2011) Robust Design for Etching Process Parameters of Hard Disk Drive Slider Fabrication Using Data Mining and Multi Response Optimization, *WSEAS Transactions On Systems And Control*, **6**(1), 15–24.

Holland, J. (1975) *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*, MIT Press, Cambridge, MA.

Huang, D., T. Allen, W. I. Notz and N. Zeng (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models, *Journal of Global Optimization*, **34**, 441–466.

Huband, S., L. Barone, L. While and P. Hinston (2005) A scalable multi-objective test problem toolkit, in *Proceedings of the 3rd International Conference on Evolutionary Multi-Criterion Optimization*, vol. 3410 of *EMO 2005*, pp. 280–294.

Hutter, F., H. H. Hoos, K. Leyton-Brown and K. P. Murphy (2009) An experimental investigation of model-based parameter optimisation: SPO and beyond, in *GECCO '09: Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, pp. 271–278, ACM, New York, NY, USA.

Hutter, F., H. H. Hoos, K. Leyton-Brown and K. P. Murphy (2010) Time-bounded sequential parameter optimization, in *Proceedings of LION-10*.

Jin, R., W. Chen and T. W. Simpson (2001) Comparative studies of metamodeling techniques under multiple modeling criteria, *Structural and Multidisciplinary Optimization*, **23**(1), 1–13.

Jin, R., W. Chen and A. Sudjianto (2002) On sequential sampling for global metamodeling in engineering design, in *ASME 2002 Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Montreal, Canada.

Jin, Y. (2005) A comprehensive survey of fitness approximation in evolutionary computation, *Soft Computing*, **9**(1), 3–12.

Jin, Y. (2011) Surrogate-assisted evolutionary computation: Recent advances and future challenges, *Swarm and Evolutionary Computation*, **1**(2), 61–70.

Johnson, M. E., L. M. Moore and D. Ylvisaker (1990) Minimax and maxmin distance design, *Journal of Statistical Planning and Inference*, **26**, 131–148.

Johnstone, I. M. and D. M. Titterington (2009) Statistical challenges of high-dimensional data., *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **367**(1906), 4237–53.

Jones, D. R. (2001) A taxonomy of global optimization methods based on response surfaces, *Journal of Global Optimization*, **21**, 345–383.

Jones, D. R., M. Schonlau and W. J. Welch (1998) Efficient global optimization of expensive black-box functions, *Journal of Global Optimization*, **13**, 455–492.

Kandil, A., K. El-Rayes and O. El-Anwar (2010) Optimization Research: Enhancing the Robustness of Large-Scale Multiobjective Optimization in Construction, *Journal of Construction Engineering and Management*, **136**(1), 17–25.

Kennard, R. W. and L. A. Stone (1969) Computer aided design of experiments, *Technometrics*, **11**(1), 137–148.

Kennedy, J. and R. Eberhart (1995) Particle swarm optimization, in *IEEE International Conference on Neural Networks*, pp. 1942–1948, Piscataway, NJ.

Khuri, A. I. and M. Conlon (1981) Simultaneous optimization of multiple responses represented by polynomial regression functions, *Technometrics*, **23**(4), 363–375.

Khuri, A. I. and J. A. Cornell (1996) *Response surfaces: designs and analyses*, Marcel Dekker, New York.

Khuri, A. I. and S. Mukhopadhyay (2010) Response surface methodology, *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(2), 128–149.

Kirkpatrick, S. (1984) Optimization by simulated annealing: Quantitative studies, *Journal of Statistical Physics*, **34**, 975–986.

Kleijnen, J. P. (2009) Kriging metamodeling in simulation: A review, *European Journal of Operational Research*, **192**(3), 707–716.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas and T. M. Cioppa (2005) State-of-the-Art Review: A User's Guide to the Brave New World of Designing Simulation Experiments, *INFORMS Journal on Computing*, **17**(3), 263–289.

Knowles, J. and E. J. Hughes (2005) Multiobjective Optimization on a Budget of 250 Evaluations, in *EMO*, pp. 176–190.

Konak, A., D. W. Coit and A. E. Smith (2006) Multi-objective optimization using genetic algorithms: A tutorial, *Reliability Engineering and System Safety*, **91**, 992–1007.

Lee, J. and P. Hajela (1996) Parallel genetic algorithms implementation for multidisciplinary rotor blade design, *Journal of Aircraft*, **33**(5), 962–969.

Li, C., F.-L. Wang, Y.-Q. Chang and Y. Liu (2009) A modified global optimization method based on surrogate model and its application in packing profile optimization of injection molding process, *The International Journal of Advanced Manufacturing Technology*.

Mai, Q. and H. Zou (2012) The Kolmogorov filter for variable screening in high-dimensional binary classification, *Biometrika*, **100**(1), 229–234.

Mansouri, S. A. (2005) A multi-objective genetic algorithm for mixed-model sequencing on JIT assembly lines, *European Journal of Operational Research*, **167**, 696–716.

McKay, N., W. J. Conover and R. J. Beckman (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21**, 239–245.

Mead, R. and D. J. Pike (1975) A review of response surface methodology from a biometric viewpoint, *Biometrics*, **31**(4), 803–851.

Meinshausen, N. and P. Bühlmann (2010) Stability selection, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**(4), 417–473.

Milborrow, S. (2009) *earth: Multivariate adaptive regression spline models*, R package version 2.3-5. **URL:** *http://CRAN.R-project.org/package=earth*.

Montgomery, D. C. (1999) Experimental Design for Product and Process Design and Development, *Journal of the Royal Statistical Society: Series D (The Statistician)*, **48**(2), 159–177.

Montgomery, D. C. (2009) *Design and analysis of experiments*, Wiley, Hoboken, NJ, 5th ed.

Myers, R. H. (1999) Response surface methodology — current status and future directions, *Journal of Quality Technology*, **31**(1), 30–44.

Myers, R. H. and W. H. Carter (1973) Response surface techniques for dual response systems, *Technometrics*, **15**, 301–317.

Myers, R. H., A. I. Khuri and G. Vining (1992) Response Surface Alternatives to the Taguchi Robust Parameter Design Approach, *The American Statistician*, **46**(2), 131–139.

Myers, R. H., A. I. Kuri and W. H. Carter (1989) Response surface methodology: 1966-1988, *Technometrics*, **31**(2), 137–157.

Myers, R. H. and D. C. Montgomery (1995) *Response surface methodology: Process and product optimization using designed experiments*, Wiley, New York.

Myers, R. H., D. C. Montgomery, G. G. Vining, C. M. Borror and S. M. Kowalski (2004) Response surface methodology: A retrospective and literature survey, *Journal of Quality Technology*, **36**(1), 53–77.

Nakagawa, T. and T. Kirikoshi (2012) Multi-Objective Robust Parameter Design Methodology Applied to Microwave Circuit, in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. II.

O'Connell, M. A. and R. D. Wolfinger (1997) Spatial regression models, response surfaces, and process optimization, *Journal of Computational and Graphical Statistics*, **6**(2), 224–241.

Paterlini, S. and T. Krink (2006) Differential evolution and particle swarm optimisation in partitional clustering, *Computational Statistics & Data Analysis*, **50**, 1220–1247.

Picheny, V., D. Ginsbourger, Y. Richet and G. Caplin (2013) Quantile-Based Optimization of Noisy Computer Experiments With Tunable Precision, *Technometrics*, **55**(1), 2–13.

Pierrot, S. (1999) Turbomachinery blade design using a navier-stokes solver and artificial neural network, *ASME Journal of Turbomachinery*, **121**(3), 326–332.

Ponweiser, W., T. Wagner, D. Biermann and M. Vincze (2008) Multiobjective Optimization on a Limited Budget of Evaluations Using Model-Assisted S-Metric Selection, in *Parallel Problem Solving from NaturePPSN X. Springer Berlin Heidelberg*, pp. 784–794.

R Development Core Team (2009) *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL: *http://www.R-project.org*.

Ren, Z., M.-T. Pham and C. S. Koh (2013) Robust Global Optimization of Electromagnetic Devices With Uncertain Design Parameters: Comparison of the Worst Case Optimization Methods and Multiobjective Optimization Approach Using Gradient Index, *IEEE Transactions on Magnetics*, **49**(2), 851–859.

Robinson, T. J., C. M. Borror and R. H. Myers (2004) Robust Parameter Design: A Review, *Quality and Reliability Engineering International*, **20**(1), 81–101.

Sacks, J., W. J. Welch, T. J. Mitchell and H. P. Wynn (1989) Design and analysis of computer experiments, *Statistical Science*, **4**(4), 409–423.

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

Santos, A. C., A. C. B. Delbem, J. B. A. London and N. G. Bretas (2010) Node-Depth Encoding and Multiobjective Evolutionary Algorithm Applied to Large-Scale Distribution System Reconfiguration, *IEEE Transactions on Power Systems*, **25**(3), 1254–1265.

Science (2011) Dealing with data (special section), in *Science*, vol. 331, pp. 639–806, American Association for the Advancement of Science.

Shan, S. and G. G. Wang (2009) Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions, *Structural and Multidisciplinary Optimization*, **41**(2), 219–241.

Sharda, B. and A. Banerjee (2013) Robust manufacturing system design using multi objective genetic algorithms, Petri nets and Bayesian uncertainty representation, *Journal of Manufacturing Systems*, **32**(2), 315–324.

Sibalija, T. V. and V. D. Majstorovic (2012) An integrated approach to optimise parameter design of multi-response processes based on Taguchi method and artificial intelligence, *Journal of Intelligent Manufacturing*, **23**(5), 1511–1528.

Simpson, T., A. J. Booker, D. Ghosh, A. A. Giunta, P. Koch and R.-J. Yang (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion, *Structural and Multidisciplinary Optimization*, **27**(5), 302–313.

Simpson, T. W., J. D. Peplinski, P. N. Koch and J. K. Allen (2001) Metamodels for Computer-based Engineering Design : Survey and recommendations, *Engineering with computers*, **17**(2), 129–150.

Simpson, T. W., V. Toropov, V. Balabanov and F. A. C. Viana (2008) Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come or Not, in *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 5, pp. 10–12.

Srinivas, N. and K. Deb (1994) Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation*, **2**, 221–248.

Ta, N. B. D., S. Zhou, W. Cai, X. Tang and R. Ayani (2011) Multi-objective zone mapping in large-scale distributed virtual environments, *Journal of Network and Computer Applications*, **34**(2), 551–561.

Taguchi, G. (1991) *Introduction to quality engineering*, New York: Unipub/Kraus International.

Tang, C. W., L. C. Chuang and H. T. Young (2012) Robust Process Design towards through-Silicon via Quality Improvement Based on Grey-Taguchi Method, *Applied Mechanics and Materials*, **217-219**, 2183–2186.

Tripathi, P. K., S. Bandyopadhyay and S. K. Pal (2007) Multi-objective particle swarm optimization with time variant inertia and acceleration coefficients, *Information Sciences*, **177**, 5033–5049.

Villanova, L., P. Falcaro, D. Carta, I. Poli, R. Hyndman, K. Smith-miles and S. Member (2010) Functionalization of Microarray Devices: Process Optimization Using a Multiobjective PSO and Multiresponse MARS Modeling, in *IEEE Congress on Evolutionary Computation (CEC2010)*, pp. 2562–2569, Barcelona.

Wang, G. G. and S. Shan (2007) Review of Metamodeling Techniques in Support of Engineering Design Optimization, *Journal of Mechanical Design*, **129**(4), 370.

Zhang, Q., W. Liu, E. Tsang and B. Virginas (2010) Expensive Multiobjective Optimization by MOEA/D with Gaussian Process Model, *IEEE Transaction on Evolutionary Computation*, **14**(3), 456–474.

Zhou, X., Y. Ma, Y. Tu and Y. Feng (2013) Ensemble of Surrogates for Dual Response Surface Modeling in Robust Parameter Design, *Quality and Reliability Engineering International*, **29**(2), 173–197.

Zhou, Z., Y. S. Ong, P. B. Nair, A. J. Keane and K. Y. Lum (2007) Combining global and local surrogate models to accelerate evolutionary optimization, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, **37**(1), 66–76.