Journal of Mathematics and Statistics 7 (1): 5-11, 2011
ISSN 1549-3644

Constrained Response Surface Optimization for a Laser Beam Welding Process

Pongchanun Luangpaiboon
Department of Industrial Engineering, Faculty of Engineering,
Industrial Statistics and Operational Research Unit,
Thammasat University, Pathumthani, 12120 Thailand

Abstract: Problem statement: On a current operating condition of a laser beam welding process in the hard disk drive industry, it has been found that shear strength of the head support and suspension assembly is slightly higher than customers' specification. This situation leads to an inspection with a large sample size and a high frequency. Shear strength is not only one quality characteristic for this assembly, but other critical physical specifications of welding diameter and depth need to be also considered. **Approach:** A hybridization strategy, based on linear and nonlinear Constrained Response Surface Optimization Methods (CRSOM), has been developed for this process refinement. The hybridization is having a provision to include both explicit constraints of influential process variables as well as implicit constraints of physical specifications. **Results:** The proposed levels of influential process variables have been successfully implemented in terms of shear strength and satisfied both welding diameter and depth specifications. **Conclusion:** The advantage of the hybridization compared with individually CRSOM is that all the data from the experiment is collected together to make a final decision. When engineering problems are large and complicated, an effective finite sequence of instructions from the hybridization can be very useful and practical in setting industrial processes such as semiconductor or automotive manufacturing systems.

Key words: Constrained optimization, shear strength, welding dept and diameter, head support and suspension assembly, Laser Beam Welding Process (LBWP), Response Surface Methodology (RSM), Evolutionary Operation (EVOP), Thailand Research Fund (TRF), Hard Disk Drive (HDD), Laser Beam Welding Process (LBWP)

INTRODUCTION

This research explains the current operating condition and engineering technologies in laser beam welding applications for a Hard Disk Drive (HDD) industry. HDD is originally introduced as efficient and reliable data storage in general computers although currently its usages are expanded into various consumer applications. HDD records and reads data by directionally magnetizing magnetic material and detecting the magnetization of the material, respectively. Typical HDD parts consist of a platter, an actuator and a recording or read-and-write head including its support and suspension assembly.

In the hard disk drive, information is stored in the form of magnetically polarized bit positions on the magnetic disk surface. Information is written to and read from the disk by a transducer which is mounted on an air bearing slider called as the head. The head will closely fly over the disk's surface. Its performance is due to the precise balance between design forces developed by the slider and a precisely controlled preload spring force to the head. While the head efficiently pitches, rolls and moves perpendicularly to the disk surface, the head could be rigidly fixed in the disk surface plane in order to avoid error reading.

In information storage systems of a magnetic disk drive the present invention mainly relates to a head support and suspension assembly or HSSA (Fig. 1). HSSA which moves and positions the recording head consists of three components. The first is a light and rigid support arm. It pivots on a servo spindle, firmly supporting the rest of the assembly as well as precisely controlling the head movement in a radial direction over the disk surface. Secondly, a load beam is attached to the rigid support arm. The proximal end of the load beam is resilient to provide a preload spring force in a direction normal to the disk surface. On a rigid portion of the load beam the preload spring force is transmitted to the distal end of the load beam where the gimbal and head are attached. The third HSSA component is a gimbal. It connects the distal end of the load beam to the head.

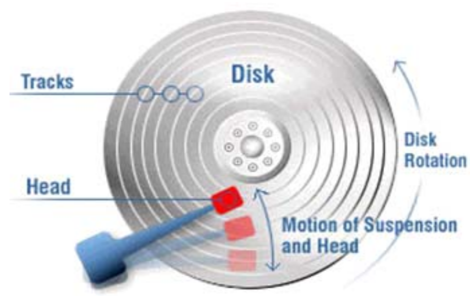

Fig. 1: Head support and suspension assembly

The gimbal is resilient in the head's pitch and roll directions to allow the slider to follow the disk topography.

In general, HSSA processes consists of etching, forming, gimbal aperture cutting, laser beam welding assembly, gram forming, separation and aqueous precision cleaning including an adjustment of gram load and static attitude. The processing variations can induce various undesirable quality measures in the HSSA. In particular, it relates to a deformation of a load beam and a gimbal or low levels of shear strength induced by a laser beam welding process used to join these components together.

MATERIALS AND METHODS

Laser Beam Welding Process (LBWP): In the welding process, laser beam with the high density is often focused and delivered to melt the work pieces and fill a material to form the weld pool that cools to become a strong joint to produce the weld (Habibi *et al*., 2009). There are two alternatives that the laser beam is delivered to the work piece. These involve the uses of hard optics and a fiber optic cable. For hard optics use, the laser beam is basically deflected and focused through only the use of mirrors and lenses. There are some practical limitations in the distance of the work piece from the laser source of this method. Work piece needs to be moved into the right position and angle to perform the weld. For the second use of a fiber optic cable, the laser energy can be focused into one end of the cable and emerge at the other end at some distance with a minor loss of energy. This use of a fiber optic cable allows for the laser beam to be precisely delivered to the needed area and even allows for movement of the focusing optics itself.

The laser power supply is capable of delivering a light pulse that consists of accurate and repeatable energy and duration. When the pulse of laser energy is focused into a small diameter spot onto the work piece, the energy density becomes quite large. The work piece absorbs the light and the focused beam vaporizes and melts some metal. As the pulse terminates the liquefied metal flows back in, solidifying and creating a small spot weld. The laser spot size or diameter ranges between 0.2 and 13.0 mm, though only smaller sizes are practically used for welding. The penetration depth of the weld is proportional to the amount of power supplied and the location of the focal point.

There are various largest advantages of a laser beam welding process when compared. This process offers minimal amount of heat added during processing. Cooling between each spot weld allows during the repeated pulsing of the beam, resulting in a very small heat affected area. This suits for welding in thin sections or products that require welding near glass-to-metal seals or electronics such as the HSSA. A process with low heat input brings greater flexibility in tooling design and materials. It almost always offers a cost advantage in both tooling and production pricing. However, successful laser welding process variables which are developed for a particular application seem impractical for others. This issue then becomes a critical matter of process control, especially in a hard disk drive industry.

In a laser beam welding process for the HSSA, some defects or unsuccessful weld results such as burning, incorrect welding depth and diameter are of the most common and detrimental issues. Not only those physical defects but unsuccessful performance results on welding as mentioned above would be considered as well. One major concern is a variation in shear strength of the HSSA. On the current operating conditions of the laser beam welding process, the process capability in shear strength is at the low level. The quality inspections are then set at high sample size and high frequency. Via the skill of the product and process engineers these unsuccessful weld and performance results on the HSSA could be affected by some compressive spring force from the fixture remaining after the laser beam welding process application with some preset levels of energy and gas flow rate. Another process variable is related to the laser beam pulse which is its sufficient intensity and width to heat through the metal.

An objective of this study is to determine the proper levels of influential process variables to maximize shear strength of the HSSA subject to satisfied physical specifications of welding diameter and depth via a hybridization of constrained response surface optimization based on the linear and nonlinear functions of variables from the laser beam welding process achieved by the regression analysis. The optimal levels of influential variables could be eventually implemented to the actual process to reduce the cost associated with quality inspections.

Constrained Response Surface Optimization Method (CRSOM): The optimal operating condition of a process response is usually determined after the sequential experiments have been conducted and a series of empirical models obtained (Al-Taweil *et al*., 2009). Different techniques can be applied in this process response optimization (AL-Marshadi, 2010). The optimization of response surfaces is different from conventional optimization techniques in various ways (Chen *et al*., 2010). Response surface optimization is an iterative procedure that is experiments performed in one set of experiments result in fitted models that indicate where to find improved levels of influential variables in the next experiment. Thus, the coefficients in the fitted model may change during the optimization process.

Moreover, the response surfaces are fitted from current experimental design points that usually contain random variability due to unknown or uncontrollable causes (Lan and Wang, 2009). If an experiment is repeated, the result will bring a different fitted response surface that may lead to different optimal levels of influential variables. Therefore, sampling variability should be concerned in the response surface optimization. It differs from the conventional optimization in which the functions to be optimized are fixed and given.

The response surface optimization is conventionally conducted in two main phases, excluding the screening experiment (Luangpaiboon, 2009). In the first phase a sequence of line searches is performed in the direction of optimal improvement. Each sequential gradient search in this steepest ascent (descent) path is terminated if there is no evidence that the direction chosen results in further improvements. The sequence of line searches is continued as long as there is no curvature effect. The second phase is performed with a second-order or quadratic polynomial regression model when there is lack of linear fit in the first phase (Ismail *et al*., 2009).

Practically, a first-order approximation will be used as a proper local search in a small region close to the initial operating conditions and far from where the process determines a curvature effect. It therefore makes sense for fitting a simple first-order or linear polynomial model from available design points. This mathematical model in forms of the search direction or path of steepest ascent (or descent) is measured to locate the new design points with a preset step length until there is no further improvement in the response. When there is a lack of fit or significant pure quadratic curvature effect, it can be implied that influential variable levels are close to where the maximum or minimum occurs. A second-order polynomial model or a canonical analysis of the surface can be used as a local approximation of the response in a small region where the optimal levels of influential variables exist.

There are some problems associated with various responses (Zandieh *et al*., 2009; Jeong *et al*., 2010; Pal and Gauri, 2010). One can be assigned as the primary or the most important response and others return to be merely secondary responses or problem constraints (Jailani *et al*., 2009; Luangpaiboon, 2010). Costa (2010) introduced the dual response surface optimization procedure focusing on the case of target value is the best. This case means keeping the mean at a specified target value whilst minimizing the standard deviation. A Constrained Response Surface Optimization Method (CRSOM) is then proposed to find optimal levels of k influential variables leading to the highest level of the primary response (y_P) and satisfying all other constraints of secondary responses (y_S). Moreover, lower and upper bounds of influential variables can be included in order to avoid solutions that extrapolate too far outside the feasible region of the experimental design points.

In order to achieve mathematical models of CRSOM, a regression analysis is used to estimate primary and associated secondary responses of $\hat{y}_p$ and $\hat{y}_S$, respectively. Both types are functions of influential variables. Linear and nonlinear programming models are then formulated with a consideration of the feasible ranges in terms of lower (LB) and upper (UB) bounds of secondary responses and influential variables (x), namely:

Optimize $\hat{y}_p$

Subject to:

$$LB \le \hat{y}_S \le UB$$
$$LB \le x \le UB$$

The new setting of influential variables will be determined to optimize the primary response, while all constraints of secondary responses and influential variables are kept in feasible ranges. Additional experiments should be performed at the optimal operating condition, estimated from the model, to confirm the satisfied levels of actual responses as expected.

Table 1: Primary and secondary responses and their feasible ranges

Classes	Responses	Feasible ranges
Primary (y_P)	Shear strength	>0.460
Secondary 1(y_{S1})	Welding depth	<0.102
Secondary 2 (y_{S2})	Welding diameter	0.18-0.26

CRSOM procedures: In this study, iterative strategies of linear and nonlinear CRSOM have shear strength (y_P) as a moving trigger or primary response whereas other physical specifications return to be merely process constraints or secondary responses (y_S) as shown in Table 1. The generation of the initial CRSOM begins with determining initial feasible design points that satisfy both implicit and explicit constraints. Implicit constraints are those that limit the values of secondary responses and explicit constraints limit the values of influential variables.

The CRSOM parameter is 10% of the significance level for tests of significance of slopes (α). The iterations replicate until the termination criteria is at the satisfaction state. Whilst continually checking termination criteria, following steps below would be carried out.

Step 1: Generate an initial set of a 2^k factorial design for k process variables located possibly near the current operating condition.

Step 2: Determine estimated regression coefficients on both primary and secondary responses by using a principle of least squares:

- Test whether there is evidence that any regression coefficient of y_P (β_i^P) and y_S (β_i^S) is different from zero at the α level of significance
- If the result is significant, formulate a linear programming model of the constrained response surface optimization method (LCRSOM) as follows

Maximize:

$$\hat{y}_P = \beta_0^P + \beta_1^P x_1 + \beta_2^P x_2 + \dots + \beta_k^P x_k$$

Subject to:

$$\begin{aligned} &LB \le \hat{y}_{S1} \text{ or } \beta_0^{S1} + \beta_1^{S1} x_1 + \beta_2^{S1} x_2 + \dots \beta_k^{P} x_k \le UB \\ &LB \le \hat{y}_{S2} \text{ or } \beta_0^{S2} + \beta_1^{S2} x_1 + \beta_2^{S2} x_2 + \dots \beta_k^{S2} x_k \le UB \\ &LB \le\ x_i \le\ UB;\ i = 1,\ 2,\ \dots,\ k \end{aligned}$$

Similarly, a nonlinear programming model of the constrained response surface optimization method (NLCRSOM) is formulated with additional estimated regression coefficients of y_P (β_{ij}^P) and y_S (β_{ij}^S).

From both mathematical models, determine the estimated design points from the CRSOM, received by the Generalized Reduced Gradient algorithm (GRG) and go to Step 3.

Step 3: If design points from the CRSOM are feasible an implementation will be carrying out via the most preferable of the mean primary responses and replace the previous operating condition by the fitted values of the CRSOM, if more preferable when compared. Check for the CRSOM termination rule or there is an evidence of curvature effect. If it does not meet, continue and go to Step 1.

RESULTS

Consider an experiment in which interest lies in examining three specifications of a laser beam welding process of the head support and suspension assembly in a hard disk drive industry. A screening experiment has identified four important process variables: laser energy (x_1), laser pulse width (x_2), gas flow rate (x_3) and fixture spring force (x_4). Changing the gas flow rate is a very simple procedure and merely makes an adjustment on a control panel while the laser beam welding process is still running. Therefore, this is the easy-to-change process variable in the experiment. Changing laser energy, laser pulse width and fixture spring force, on the other hand, requires the precise adjustment. Thus, these three variables are hard-to-change process variables. The (low, high) levels in the natural variables for x_1, x_2, x_3 and x_4 are coded as (295, 305), (450, 600), (20, 40) and (3, 7), respectively.

A preliminary study via the normal probability plot of effects allows experiments to be more efficient and use fewer runs. It aims to reduce many candidate process variables to a relatively few potentially important variables that influence the responses. The results showed that influential variables of x_1, x_2 and x_4 affect the primary response of shear strength (Fig. 2). However, only process variables x_1 and x_2 are influential to the secondary responses of welding depth and diameter at the approximate 95% confidence interval throughout. Obtained information is used to move through the experimental region in an attempt to get closer to the process optimum.

A hybridization of linear and nonlinear constrained response surface optimization is applied to derive the new expected operating design points based on the group structure of the design point arrays. The first iteration provided three cases of estimated levels of influential variables.

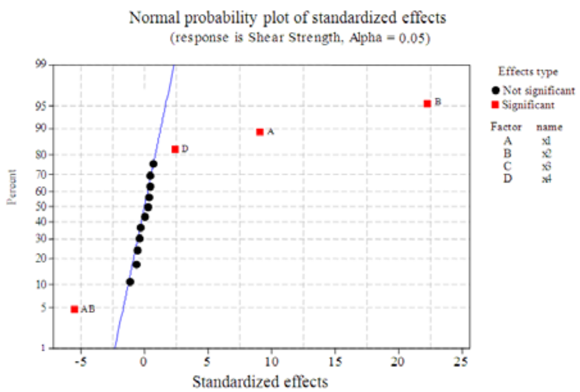

Fig. 2: Normal probability plot of effects on a primary response of shear strength

Fig. 3: Melted work piece from the LBWP via the LCRSOM

These design points will be used to analyze the highest average primary responses whilst satisfying all constraints of secondary responses and influential variables. In each of these settings, some applicable replicates are run. The average of these runs will serve as the response value for the preferable alternative. On the LCRSOM the estimated setting was impractical whereas the nonlinear model gave the new setting of influential variable with the highest average responses. The variable setting from the NLCRSOM could be used as the new operating condition of a laser beam welding process for any given periods of follow-up experimentation. Continue performing the new setting until the next iteration of experimentation is determined.

During the second iteration, changing the levels of the laser energy, laser pulse width and fixture spring force according to the estimated results from the LCRSOM, required the process to be stopped because the melted work piece occurred (Fig. 3). Typically, there were some improvements before the response started to drop off. For useful information the previously determined settings for two iterations could be used for any follow-up experimentation of the LCRSOM or as alternative of the NLCRSOM. The values of the process variable levels via the NLCRSOM led to the better shear strength, welding diameter and depth when compared.

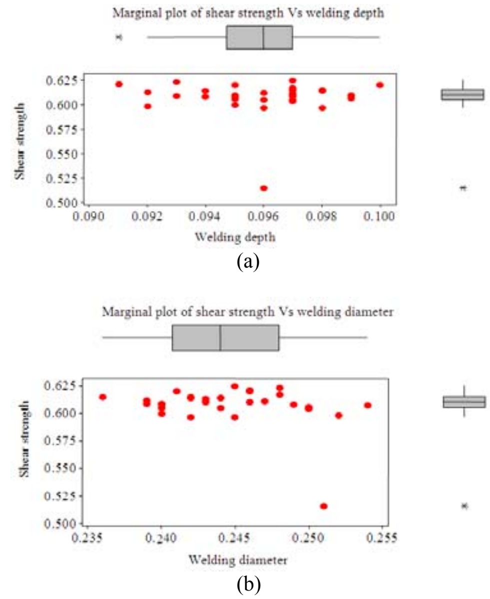

Fig. 4: Marginal plot of primary response against secondary responses of welding depth (a) and welding diameter (b)

Table 2: Regression coefficients and their P-Values categorized by responses for the nonlinear CRSOM

Influential Variables	Constant	x_1	x_2	x_1x_2
Shear Strength				
Coefficient	-13.260	0.043602	0.022790	-0.00007161
P-Value		0.000	0.000	0.000
Welding Depth				
Coefficient	1.819	-0.006241	-0.004651	0.00001615
P-Value		0.000	0.000	0.018
Welding Diameter				
Coefficient	-2.019	0.007177	0.0028841	-0.00000911
P-Value		0.000	0.008	0.025

In the third iteration, more settings of factorial design points were used, all with the centre from the previous operating condition. The responses of all the settings were used to determine the regression coefficients for formulating the NLCRSOM (Table 2), with the preset level of a compressive spring force (x_4) at 8.

Consider the nonlinear model of the CRSOM fit with the coefficients from Table 2. Then, the new estimated design points, achieved by GRG, was x_1 = 314.97, x_2 = 554.74 and x_4 = 8. The LCRSOM and NLCRSOM were repeated and the results appeared in Table 3.

After an implementation of the hybridization, it has been found that the results on shear strength seem to be better and remaining secondary responses are still within their specifications as shown in forms of a scatter plot with box plots in the margins called as the marginal plots (Fig. 4).

Table 3: Iterative experiments on the LBWP via a hybridisation of LCRSOM and NLCRSOM, where x_3 fixed at 30

Iteration	Alternatives	(x_1, x_2, x_4, Actual y_P)
1	LCRSOM	(84.75, 2837.15, 8.00, -)
	NLCRSOM	(314.50, 545.92, 8.00, 0.607)
2	LCRSOM	(303.77, 663.13, 8.00, 0.627)
	NLCRSOM	(315.12, 535.42, 8.00, 0.611)
3	LCRSOM	(305.03, 600.25, 8.00, 0.608)
	NLCRSOM	(314.97, 554.74, 8.00, 0.625)

DISCUSSION

It is important to keep in mind the nature in continually determining the iterations in a context of evolutionary operations. The sequential procedures of the proposed strategy needed to further improve the process. However, the investigation in this particular example terminates with the preset number of iterations of the hybridization of linear and nonlinear CRSOM is three. Moreover, from a practical point of view if the largest standardized regression coefficient belongs to the most influential variables with satisfied levels of P-Values, the NLCRSOM seems like an obvious alternative to be chosen. This would allow the experimenter to make the great step length in the most influential variables toward the optimum and then minor refinements in the less influential variables.

In contrast to this, if the largest standardized regression coefficient belongs to the most influential variables with undesirable levels of P-Values, the LCRSOM seems to be better because its movement should provide the quicker step towards the optimum. Each method has its own advantages and disadvantages. The best method will rely on the particular experimental circumstances, the starting design points used, the nature or shape of the true response surface, experience and results from previous experiments. However, in some iteration it is possible to combine or switch between these two methods.

When the main and interaction effects are considered, they indicate some relationships of a primary response of shear strength and other secondary responses on influential variables (Table 4). Shear strength and welding diameter have similar effects on influential variables. Both responses increase as moving from the low to the high levels of influential variables. In contrast to this, when moving from the low to the high levels of influential variables shear strength increases, but welding depth decreases.

Despite the shortcomings of conventional response surfaces of influential variables and each process response, the relationship of all responses is useful for our research. Experimental results from three iterations are then used to form a typical contour plot of three responses (Fig. 5). This enables our exploration of the primary response that similarly influences other physical measures of secondary responses as previously described on Table 3.

Table 4: Relationships between responses and influential variables

Responses		Main effects x_1, x_2, x_4	Interaction effects x_1 x_2
Primary	Shear Strength	+, +, +	-
Secondary	Welding Depth	-, -	+
	Welding Diameter	+, +	-

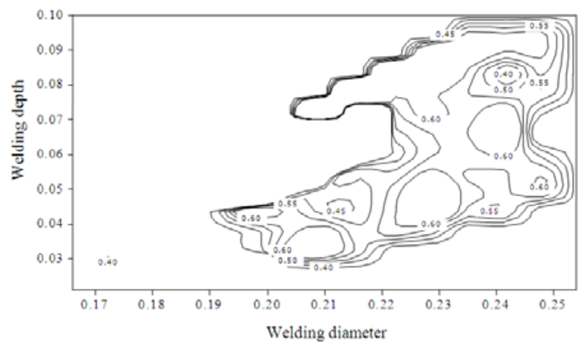

Fig. 5: Contour plot of primary against secondary responses

CONCLUSION

Various sequential tools in terms of response surface methods are widely used in industrial optimizations to determine the most preferable conditions of influential variables. Nowadays many of these industrial optimizations involve one or more responses and some restrictions on the feasible ranges of influential variables. Changing the level of one influential variable for one response may affect other variables or other responses on the tool. Therefore, it is important to efficiently collect and analyze all of the data at one time. In a case of no dominant response to be classified, the desirability function should be applied. Almost inevitably, some effective trade-off has to be made in order to determine process operating conditions that are satisfactory for all responses. In contrast to this, when one response can be clearly chosen as the primary and bounds or targets can be defined on all other responses, a programming model based approach can be taken.

Under the latter circumstances, a hybridization of linear and nonlinear constrained response surface optimization methods are proposed in this research. A preliminary study is required to indicate the dependence on the responses and process variables. An investigation of the performance of the hybridization is then carried out to enhance the possibility of moving towards the optimum of the laser beam welding process. There is no simple selection to choose between them in a constrained problem manner, where, by the optimal condition, the preferable method would lead to the most rapid movement to the feasible region of the optimum. Two alternatives of mathematical models of CRSOM are applied to determine the estimated design point to me*et al*l the constraints with the most preferred level of the primary response. However, applying linear and nonlinear programming models on CRSOM to this process resulted in little significant difference in terms of the mean actual yields received.

As stated earlier, the proposed strategy in this research was restricted to three influential variables and three responses and the error standard deviation was at higher levels. The improvement via these sequential procedures seems to be slightly slow. Other stochastic subprocedures from heuristics such as ant colony optimization or harmony search algorithms could be applied to this strategy to increase its performance, especially in terms of speed of convergence.

ACKNOWLEDGMENT

This study was supported by the Thailand Research Fund (TRF), the National Research Council of Thailand (NRCT) and the Commission on Higher Education of Thailand.

REFERENCES

Al-Taweil, H.I., M.B. Osman, A.A. Hamid and W.M.W. Yusoff, 2009. Optimizing of trichoderma viride cultivation in submerged state fermentation. Am. J. Applied Sci., 6: 1284-1288. DOI: 10.3844/ajassp.2009.1284.1288

AL-Marshadi, A.H., 2010. Simulation study for evaluating a new three stages procedure for selecting the right multivariate regression model. Am. J. Biostat., 1: 32-41. DOI: 10.3844/amjbsp.2010.32.41

Chen, G., L. Tang and Z. Li, 2010. Parameter optimisation of human-simulated intelligent controller for a cart-double pendulum system. Int. J. Modelling Identification Control, 10: 194-201. DOI: 10.1504/IJMIC.2010.034571

Costa, N.R.P., 2010. Simultaneous optimization of mean and standard deviation. Quality Eng., 22: 140-149. DOI: 10.1080/08982110903394205

Habibi, M., H. Shahmohammadi, V. Taraghi, S.D. Safari and B. Arezoo, 2009. A prototype two-axis laser scanning system used in stereolithography apparatus with new algorithms for computerized model slicing. Am. J. Applied Sci., 6: 1701-1707. DOI: 10.3844/ajassp.2009.1701.1707

Ismail, Z., A. Yahya and A. Shabri, 2009. Forecasting gold prices using multiple linear regression method. Am. J. Applied Sci., 6: 1509-1514. DOI: 10.3844/ajassp.2009.1509.1514

Jailani, H.S., A. Rajadurai, B. Mohan, A.S. Kumar and T. Sornakumar , 2009. Multi-response optimization of sintering parameters of Al-Si alloy/fly ash composite using Taguchi method and grey relational analysis. Int. J. Adv. Manufact. Technol., 45: 362-369. DOI: 10.1007/s00170-009-1973-3

Jeong, I.J., K.J. Kim and D.K.J. Lin, 2010. Bayesian analysis for weighted mean-squared error in dual response surface optimization. Quality Reliabil. Eng. Int., 26: 417-430. DOI: 10.1002/qre.1058

Lan, T.S. and M.Y., Wang, 2009. Competitive parameter optimization of multi-quality CNC turning. Int. J. Adv. Manuf. Technol., 41: 820-826. DOI: 10.1007/s00170-008-1495-4

Luangpaiboon, P., 2009. Hybridizations of simulated annealing and modified simplex algorithms on a path of steepest ascent with multi-response for optimal parameter settings of ACO. AIP Conf. Proc., 1174: 94-108. DOI:10.1063/1.3256264

Luangpaiboon, P., 2010. Improving an electrostatic powder coating process via signal to noise response surface. Am. J. Applied Sci., 7: 1521-1527. DOI: 10.3844/ajassp.2010.1521.1527

Pal, S. and S.K. Gauri, 2010. Multi-response optimization using Multiple Regression-Based Weighted Signal-to-Noise Ratio (MRWSN). Qual. Eng., 22: 336-350. DOI: 10.1080/08982112.2010.495368

Zandieh, M., M. Amiri, B. Vahdani and R. Soltani, 2009. A robust parameter design for multi-response problems. J. Comput. Applied Math., 230: 463-476. DOI: 10.1016/j.cam.2008.12.019