Augmented Approach to Desirability Function Based on Principal Component Analysis

Yuvarani Thorisingam[1], Mohd Shafie Mustafa[1,2,a]
[1]Institute for Mathematical Research, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia, [2]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
Corresponding Author Email: shafie@upm.edu.my

Abstract
The desirability function approach is commonly used in industry to tackle multiple response optimization problems. The shortcoming of this approach is that the variability and correlated in each predicted response are ignored. It is now evident that the actual response may fall outside the acceptable region even though the predicted response at the optimal solution has a high overall desirability score. An augmented approach to the desirability function (AADF) and the Principal Component Analysis (PCA) is put forward to rectify this problem. This paper will discuss how the two methodologies have been used together where the goal is to determine the final optimal factor/level combination when several responses are to be optimized. Additionally, in this work optimization of multiple correlated responses was studied and AADF model was proposed based on PCA to optimize correlated multiple response problems. The proposed method is also demonstrated by numerical example from the literature to confirm the efficiencies.
Keywords: Response Surface Designs, Optimal Design, Multivariate Analysis, Principal Component Analysis

Introduction
      As approaching 4th Industry Revolution, optimization plays big role in determining the survival of industry by reducing large number of variables from a huge dataset or experiment. Big data is providing rich information about every process and product and the quality improvement has become a big player in many technical fields. It is often expensive to manufacture a poorly designed product. In process and product optimization, a common problem is to determine the best optimal performance measure of the product or process. Usually there is more than one parameter of the product that must in some sense be simultaneously considered. Essentially, this becomes a problem in the multiple responses, each of which depends upon a set of factors (Natabirwa et al., 2018).
      The handling process of huge data sets is challenging to most of the researchers. Thus, PCA is used to reduce the dimensionality of a data set consisting of large number of interrelated variables, while retaining as much as possible of the variation present in the data

set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all the original variables (Jolliffe, 2002). Like other optimization work in Response Surface Methodology, RSM the desirability function is widely used to solve the simultaneous optimization problem (Ribeiro, 2010). The desirability function work smoothly even with many responses are involved in the optimization process. According to Chen et al. (2012), augmented desirability function offers multiple compromise solutions and greater flexibility. This is due to the ability of incorporate various types of information or optimality criteria with the relative weight.

To date, many studies conducted on optimize multiple responses. However, there is no study has been conducted on optimizing multiple response by retaining the maximum quality characteristics of study while take into consideration on variability. Hence, this inspire us to formulate an augmented approach to desirability function (AADF) by using principal component analysis (PCA) to reduce the variation among responses.

Literature Review

Numerous responses in data set are not easy to analyse. The dispersion matrix of a data set is too large to study and interpret. Thus, it is necessary the data to undergo data reduction process by PCA according to (Wang et al., 2013). Apart from data reduction, the quality improvement is a very important factor in a product or process development. One of the methods that popular in quality improvement is RSM. In stated (Kwak, 2005), RSM successfully minimized geometric error in the process of grinding process.

Generally, the number of response is greater than three in a data set. In contrast, RSM works better with three variables in data set. In these situations, PCA can be used as data reduction method to reduce the sensory dimensions to a more manageable size according to Rossi (2001). Desirability function is required to obtain optimum condition or level of process. In 1965, desirability function was introduced by Harrington using geometric mean function. Meanwhile, in 1994, weighted geometric mean was proposed by Derringer. The goal of both methods is to determine the factor settings that maximize the weighted geometric mean of the individual desirability functions. However, modification of desirability function is widely used in many real problems which suggested by (Derringer and Suich, 1980). This method considers an objective function initially which transforms the existing values in to a scale free value called desirability.

In 2012, Chen et al. minimized prediction variation in desirability function. The desirability function ignores the variability in each predicted response. Therefore, actual response at the optimal solution might be fall in rejection region. This situation creates higher chances to the actual response at the optimal solution will not be acceptable.

Methodology

The desirability approach to simultaneously optimizing multiple equations was proposed by Harrington (1965). Desirability is an objective function that ranges from zero outside of the limits to one at the goal. The numerical optimization finds a point that maximizes the desirability function. For several responses and factors, all goals get combined into one desirability function. The desirability value is completely dependent on how closely the lower and upper limits are set relative to the actual optimum. Transform each of the y predicted response to an individual desirability function d_i where $0 \leq d_i \leq 1$.

Harrington (1965) transformed y predicted to desirability function using exponential function,

a. one-sided transformation, $d_i = -\exp\left(-|\widehat{y_i}|^r\right)$

b. two-sided transformation, $d_i = \exp(-|\widehat{y_i}|^r)$

where r is an user selected shape parameter based on experimenter opinion.

Overall desirability function, $D = (d_1^{\square} \, d_2^{\square} \, \dots \, d_m^{\square})^{1/m}$ calculated using geometric mean. Chen et al. (2012) used individual desirability of smaller the better (STB) to calculate desirability function by transforming standard deviation,

$$d_i = \begin{cases} 1 & for\ \widehat{y_i} \le y_i^{min} \\ (\frac{y_i^{max}-\widehat{y_i}}{y_i^{max}-y_i^{min}})^r & for\ y_i^{min} < \widehat{y_i} < y_i^{max} \\ 0 & for\ \widehat{y_i} \ge y_i^{max} \end{cases}$$

Then all the individual desirability function are combined as function S using geometric mean,

$$S = (d_{s1}^{\square} \, d_{s2}^{\square} \, \dots \, d_{sm}^{\square})^{1/m}$$

Using Harrington's overall desirability D, the augmented approached defined as

$$DS_\lambda = D^\lambda S^{1-\lambda^{\square}}$$

where $0 < \lambda \le 1$ is a user-selected weight that reflects the relative importance of optimizing D and S. When $\lambda = 1$, DS_λ reduces to D, which is identical to Harrington's desirability function.

The principal components obtained from the PCA treated as responses by RSMRSM is created with two main objectives. First goal of RSM is to adequate functional relationship between a response of interest and several associated control variables. The second goal is to obtain the optimum settings of the variables that yield maximum or minimum response. We assume that the i th response y_i may be approximated by a linear regression model,

$$y_i = x\beta_i + \varepsilon_i, \quad i = 1,2,\dots,m$$

where β_i is a vector and $\varepsilon_i \sim N(0, \sigma_i^2)$.

Each predicted response at the point x has variance,

$$sd(y) = \sigma_i^2 v_i(x)$$

The narrower prediction interval of the response gives more precise information of the predicted response by retaining the quality characteristics.

Result and Discussion

In order to demonstrate the application of the proposed method, the real dataset from Hu et al (2008) is considered in Table 1. In this example, there are three response variables (y1, y2, y3) and three design variables (x1, x2, x3). The target responses of y1, y2 and y3 under the range of $[y_i^{min}, y_i^{max}] = [93,100]$ according to experimenter. The experimenter set the weight, Λ=0.9 to find the optimal solution of x_*. We observe that when $\lambda \le 0.9$, the values of S are larger than the one at $\lambda =1$. This indicates the standard deviation of y obtained from our augmented approach are smaller.

The experiment conducted using central composite design and assumed 2nd order polynomial model for the responses. The correlated mean and standard deviations of the responses are transformed into uncorrelated component through PCA. The eigenvalues and eigenvectors used for the PCA is shown in the Table 3 and Table 4.

The target value obtained by the RSM is pass through the principal component model to attain the desirability function. The obtained value is compared with the previous

approaches by Harrington and Chen. The comparison between previous approaches is shown in Table 4.

Table 1
The value of variable, responses and principal components.

x1	x2	x3	x4	y1	y2	y3	PC1	PC2	PC3
-1	-1	-1	-1	87.43	81.79	84.97	-1.902	-0.30427	-0.12041
-1	-1	-1	1	85.74	81.26	83.27	-2.47532	-0.03777	-0.25695
-1	-1	1	-1	87.49	84.41	90.09	-0.90144	0.175811	0.549911
-1	-1	1	1	84.91	84.1	85.7	-1.9748	0.561205	-0.04197
-1	1	-1	-1	91.16	89.4	92.82	0.613153	-0.0321	0.270745
-1	1	-1	1	88.36	90.94	92.25	0.231388	0.762016	0.209074
-1	1	1	-1	92.58	90.2	93.39	1.012316	-0.25387	0.190062
-1	1	1	1	88.08	88.43	91.25	-0.21873	0.516918	0.304764
1	-1	-1	-1	87.3	88.15	86.21	-1.08131	0.458585	-0.54105
1	-1	-1	1	84.17	86.61	85.58	-1.85018	0.997471	-0.27059
1	-1	1	-1	90.49	91.71	91.08	0.502111	0.298838	-0.24163
1	-1	1	1	87.35	89.46	89.31	-0.5023	0.716491	-0.10754
1	1	-1	-1	93.94	90.83	93.54	1.324996	-0.49738	0.053734
1	1	-1	1	87.34	90.3	92.28	-0.00084	0.932657	0.355018
1	1	1	-1	94.29	92.33	94.7	1.702809	-0.37004	0.087224
1	1	1	1	93.25	93.35	92.82	1.375171	-0.09673	-0.29044
-2	0	0	0	90.2	86.93	90.32	-0.15495	-0.1769	0.134439
2	0	0	0	92.26	92.48	92.91	1.132095	0.044793	-0.11175
0	-2	0	0	88.64	83.49	89.68	-0.86483	-0.20799	0.486
0	2	0	0	94.23	94.37	95.43	2.009414	-0.10722	0.015039
0	0	-2	0	88.95	82.68	85.65	-1.46143	-0.53514	-0.19735
0	0	2	0	93.53	93.33	93.81	1.557734	-0.12288	-0.12573
0	0	0	-2	86.07	74.08	81.25	-3.45804	-0.96759	0.090007
0	0	0	2	84.72	74.35	80.52	-3.75538	-0.65428	0.025126
0	0	0	0	92.25	91.34	92.69	0.979786	-0.08408	-0.03352
0	0	0	0	94.02	92.18	93.29	1.445368	-0.38196	-0.13821
0	0	0	0	93.21	92.24	93.72	1.377448	-0.16867	-0.00615
0	0	0	0	93.78	92.57	94.4	1.601617	-0.23792	0.043932
0	0	0	0	93.87	94.91	94.98	1.943702	0.015771	-0.09796
0	0	0	0	94.39	94.02	93.95	1.792442	-0.24375	-0.23383

Table 2

The eigenvalue from principal component analysis.

Components	Eigenvalues	Proportion	Cumulative Percentage
PC_1	54.891	0.921	0.921
PC_2	3.552	0.060	0.981
PC_3	1.139	0.019	1.000

Table 3

The eigenvector for each principal component analysis.

Responses	PC_1	PC_2	PC_3
$y1$	0.403	-0.814	-0.419
$y2$	0.724	0.564	-0.398
$y3$	0.560	-0.143	0.816

Table 4 compares the optimization results from Harrigton's desirability function and Chen Augmented approach. The variance and standard deviation of each predicted responses shows minimal with 0.71 for proposed method compared with the other two methods. This indicates that the proposed method able to calculate the narrower prediction and this proof that optimal solution are predicted more precisely. This data shows that method used from the proposed method provide comparable optimization output from the other two methods.

Table 4
The result of comparison study.

Optimization Results	Response	$\widehat{y_\iota}$	d_i	sd($\widehat{y_\iota}$)	90% Prediction interval	Length
Harrington's desirability $\Lambda=1$, $\quad x_* = (0.82, 1.88, 0.61, -0.1)$ $D=0.8089$	y1	96.24	0.79	1.16	(93.12,99.37)	6.25
	y2	95.81	0.76	2.06	(90.27, 101.34)	11.07
	y3	97.52	0.88	1.27	(94.09,100.94)	6.85
Chen Augmented Approach $\Lambda=0.9$, $\quad x_* = (0.6, 1.19, 0.5, -0.09)$ $D=0.7839$, $\quad$ S=0.44, DS=0.74	y1	95.94	0.77	0.65	(93.31,98.56)	5.25
	y2	95.77	0.75	1.15	(91.12, 100.42)	9.30
	y3	96.70	0.83	0.71	(93.82,99.58)	5.76
Proposed Method with PCA $\Lambda=0.9$, $\quad x_* = (0.6, 1.19, 0.5, -0.09)$ $D=0.746$, $\quad$ S=0.41, DS=0.72	y1	95.43	0.71	0.36	(94.09,98.74)	4.65
	y2	95.66	0.74	0.42	(92.32, 99.73)	7.41
	y3	96.02	0.79	0.27	(93.25,98.84)	5.59

Conclusion

In this article, we proposed an augmented approach to the desirability function (AADF) to determine the final optimal factor/level combination when several responses are to be optimized by reducing variation. The Harrington's desirability and Chen Augmented Approach methods are not effective as it shows broad prediction interval compared with AADF. The proposed method achieved less variation compared with classical methods to reduce highly correlated data by retaining the quality characteristics. Thus, AADF is highly recommended to optimize multiple response with minor variation and narrower prediction internal.

Reference

Wang, J., & Wang, X. (2013). A wear particle identification method by combining principal component analysis and grey relational analysis. *Wear, 304*(1-2), 96-102.

Kwak, J. (2005). Application of Taguchi and response surface methodologies for geometric error in surface grinding process. *International Journal of Machine Tools and Manufacture, 45*(3), 327-334.

Rossi, F. (2001). Blending response surface methodology and principal components analysis to match a target product. *Food Quality and Preference, 12*(5-7), 457-465.

Derringer, G., Suich, R., 1980. Simultaneous optimization of several response variables. *Journal of Quality Technology, 12* (4), 214–219.

Chen, H., Wong, W. K., & Xu, H. (2012). An augmented approach to the desirability function. *Journal of Applied Statistics, 39*(3), 599-613.

Chen, H., Xu, H., & Wong, W. K. (2012). Balancing Location and Dispersion Effects for Multiple Responses. *Quality and Reliability Engineering International, 29*(4), 607-615.

Jeong, I., Kim, K. (2009). An interactive desirability function method to multiresponse optimization. *European Journal of Operational Research, 195(*2), 412-426.

Pal, S., Gauri, S. K. (2018). A desirability functions-based approach for simultaneous optimization of quantitative and ordinal response variables in industrial processes. *International Journal of Engineering, Science and Technology, 10*(1), 76.

Majumder, H., Paul, T., Dey, V., Dutta, P., Saha, A. (2017). Use of PCA-grey analysis and RSM to model cutting time and surface finish of Inconel 800 during wire electro discharge cutting. *Measurement, 107*, 19-30.

Wonggasem, K., Wagner, T., Trautmann, H., Biermann, D., Weihs, C. (2013). Multi-objective Optimization of Hard Turning of AISI 6150 Using PCA-based Desirability Index for Correlated Objectives. *Procedia CIRP, 12*, 13-18.

Ribeiro, J., Tefilo, R., Augusto, F., Ferreira, M. (2010). Simultaneous optimization of the microextraction of coffee volatiles using response surface methodology and principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 102*(1), 45-52.

Pearson, K., (1901). On line and planes of closest fit to systems of points in space. *Philosophical Magazine 6*: 559-572.

Hotelling, H., (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational of Psychology 24:* 417-441.

Hu, Z., Cai, M., & Liang, H.-H. (2008). Desirability function approach for the optimization of microwave-assisted extraction of Saikosaponins from radix bupleuri. *Separation and Purification Technology*, *61*(3), 266–275.

Castillo, E. D., Montgomery, D. C., Mccarville, D. R. (1996). Modified Desirability Functions for Multiple Response Optimization. *Journal of Quality Technology, 28*(3), 337-345.