

# Quantile Plots of the Prediction Variance for Partially Replicated Central Composite Design

Ngozi C. Umelo-Ibemere<sup>1</sup> & Harrison O. Amuji<sup>1</sup>

<sup>1</sup> Federal University of Technology, Owerri

Correspondence: Ngozi C. Umelo-Ibemere, Department Of Statistics, Federal University Of Technology, Owerri.  
E-mail: ngorismain@yahoo.com

Received: January 3, 2015 Accepted: January 21, 2015 Online Published: March 31, 2015

doi:10.5539/ijsp.v4n2p55 URL: <http://dx.doi.org/10.5539/ijsp.v4n2p55>

## Abstract

Sometimes, it is not feasible to fully replicate the experimental units. When this happen, there is need for optimal replication of the experimental units to avoid bias. The prediction variance of two variations of the partially replicated central composite design (replicated cube plus one star and one cube plus replicated star) are compared using the quantile plots. These plots provide information about the prediction variance distribution on a sphere for comprehensive evaluation of the quality of the prediction variance. For face-centred ( $\alpha = 1$ ) and rotatable ( $\alpha = F^{\frac{1}{4}}$ ) central composite designs, the prediction variance of the one cube plus replicated star perform better than the replicated cube plus one star. Unlike the replicated cube plus one star, the quantile plots of the scaled prediction variance of the one cube plus replicated star depict near rotatability.

**Keywords:** central composite design, partial replication, prediction variance, quantile plots, rotatability

## 1. Introduction

Replication is one of the three basic principles of experimental design. Contextually, it implies independent repetition of each combination of factor levels. It helps to obtain accurate estimates of experimental errors and the effects of the input variables on the response variable(s). The analysis of such experiments have been described by many authors; see, for example, Montgomery (2012), Dean and Voss (1999), Atkinson and Donev (1992). Sometimes, it is not feasible to replicate each unique combination of factor levels. The number of experimental runs required can outgrow the resources of most experiments. An estimate of the experimental error can still be obtained by replicating only some points in an experimental design. This is referred to as partial replication. However, possible bias that may be introduced by selectively replicating partially must be addressed. Replicating only the runs that are most easily repeated may lead to inadvertently choosing the combinations of factor levels that may produce very little (or too much) random variability. This of course, poses a problem for the experimenter in choosing the points to replicate and the points not to replicate in a given design. In this work we are interested in central composite design (ccd), which is perhaps the most popular class of second order designs that is practically useful. A ccd comprises of a factorial part,  $F$  consisting of units of  $2^{k-q}$  ( $q \geq 0$ ) of at least resolution V (a situation where the main effects and two-factor interactions are not aliased with any other main effects or two-factor interactions) with each point replicated  $n_F$  times, which is usually called the cube. The levels of the factors are coded  $(\pm 1, \pm 1)$ ; an axial part consisting of  $2k$  units on the axis of each factor at a distance,  $\alpha$ , from the centre of the design,  $[(\pm\alpha, 0), (0, \pm\alpha)]$  usually called the star, with each point replicated  $n_\alpha$  times and  $n_0$  replication of the centre points,  $(0, 0, \dots, 0)$ ; all of which give a total of:  $N = n_F 2^{k-q} + n_\alpha 2k + n_0$ , (Draper, 1982).  $\alpha$  values are chosen based on some design restrictions such as rotatability, slope-rotatability; see for example, Hader and Park (1978). Often in design comparison, single value design optimality criteria are used, see, for example Lucas (1976). Chigbu and Ohaegbulam (2011) and Nduka and Chigbu (2014) compared two variations of the partially replicated orthogonal and rotatable ccd using  $D$ - optimality criteria and Schur ordering for 2 to 5 input variables. The variations are the replicated cube plus one star and the replicated star plus one cube. Their results show that replicated cubes are preferred to replicated stars. Also, designs can be compared by evaluating the prediction variances throughout the experimental region. In comparing the prediction capabilities of some response surface designs, the Variance Dispersion Graph (VDG) proposed by Giovannitti-Jensen and Myers (1989) has been used; see, for example, Borkowski(1995), Li, Liang, Borrer, Anderson-Cook and Montgomery (2009). Ukaegbu and Chigbu (2015) used the VDG and Fraction

of Design Space (FDS) to compare the prediction variances of partially replicated ccd in cuboidal region and conclude that the replicated star is economically feasible than the replicated cube. VDG does not provide enough information on the effect of design on the prediction variance, hence, Khuri, Kim and Um (1996) proposed a method for describing the distribution of the predicted variances on a given sphere in terms of its quantiles. Unlike the VDG, the quantile plots provide information not limited to the maxima and minima of distribution of the predicted variances. In this work, we shall use the quantile plots to compare two variations of a partially replicated  $N$ -point ccd. The variations are the replicated cube plus one star and the one cube plus replicated star.

**2. Method**

*2.1 Prediction Variance*

Consider a second order model in  $k$  variables of interest,  $x_1, x_2, \dots, x_k$

$$y(x) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \sum_j \beta_{ij} x_i x_j + e \tag{1}$$

which can be written as

$$y(\mathbf{x}) = f'(\mathbf{x})\beta \tag{2}$$

The fitted model of ( 1 ) can be expressed as

$$\hat{y}(x) = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i + \sum_{i=1}^k \hat{\beta}_{ii} x_i^2 + \sum_{i < j} \sum_j \hat{\beta}_{ij} x_i x_j \tag{3}$$

or

$$\hat{y}(\mathbf{x}) = f'(\mathbf{x})\hat{\beta} \tag{4}$$

where ,  $\mathbf{x} = (x_1, x_2, \dots, x_k)$ ,  $f'(\mathbf{x}) = (1, x_1, x_2, \dots, x_k, x_1^2, x_2^2, \dots, x_k^2, x_1 x_2, x_1 x_3, \dots, x_{k-1} x_k)$  and  $\beta$  is an  $m \times 1$  vector of constant coefficients. The least square estimator of  $\beta$  is given by

$$\hat{\beta} = (X'X)^{-1}X'y \tag{5}$$

$X$  is an  $N \times m$  matrix of rank  $m$  and  $y$  is the vector of response values. The variance-covariance matrix of  $\beta$  is

$$Var(\hat{\beta}) = (X'X)^{-1}\sigma^2 \tag{6}$$

$\sigma^2$  denotes the variance of  $e$ . The variance of the predicted value or the prediction variance is

$$Var[\hat{y}(\mathbf{x})] = \sigma^2 f'(\mathbf{x})(X'X)^{-1} f(\mathbf{x}) \tag{7}$$

In comparing design of different sizes, it is convenient to scale the prediction variance, see Atkinson and Donev (1992), Myers, Montgomery and Anderson-Cook (2009). Hence the scaled prediction variance (SPV) is expressed as

$$\eta(\mathbf{x}) = \frac{N}{\sigma^2} Var[\hat{y}(\mathbf{x})] \tag{8}$$

$$\eta(\mathbf{x}) = N f'(\mathbf{x})(X'X)^{-1} f(\mathbf{x}) \tag{9}$$

If the value of  $\eta(\mathbf{x})$ , which of course depends on the points,  $x_1, x_2, \dots, x_k$ , is a function, only of the distance from the centre of the design, the design is said to be rotatable. In other words, a rotatable design is one for which the quality of the estimator,  $\hat{y}$  and  $Var[\hat{y}(\mathbf{x})]$  are the same for two points that are of the same distance from the design centre. When we have no prior knowledge of the location of the optimum in an experimental region and the direction to take for improvement, it makes sense to use a design that provides equal precision of estimation in all directions. For a rotatable central composite design,  $\alpha = F^{\frac{1}{4}}$ , ( see Myers, 1971).

## 2.2 Review of the Quantiles of SPV

We shall briefly review the method of evaluating the quantiles of SPV of response surface designs as proposed by Khuri, Kim & Um (1996). Estimates of the quantiles of  $\eta(\mathbf{x})$  in ( 8 ) on a sphere,  $S(r)$

$$S(r) = \left\{ \mathbf{x} : \sum_{i=1}^k x_i^2 \right\} \quad (10)$$

of radius  $r$  inside a region of interest,  $R$  can be obtained by randomly selecting a large number of points on  $S(r)$ , then obtaining the value of  $\eta(\mathbf{x})$  at each of these points. The random selection of points is achieved by using spherical coordinates. Any point  $\mathbf{x} = (x_1, x_2, \dots, x_k)$  on  $S(r)$  can be represented using  $k - 1$  independent spherical coordinates  $\theta_1, \theta_2, \dots, \theta_{k-1}$  such that

$$\begin{aligned} x_1 &= r \cos \theta_1, \\ x_2 &= r \sin \theta_1, \cos \theta_2, \\ x_3 &= r \sin \theta_1, \sin \theta_2, \cos \theta_3, \\ &\vdots \\ x_{k-2} &= r \sin \theta_1, \sin \theta_2, \dots, \sin \theta_{k-3} \cos \theta_{k-2}, \\ x_{k-1} &= r \sin \theta_1, \sin \theta_2, \dots, \sin \theta_{k-3} \sin \theta_{k-2} \cos \theta_{k-1}, \\ x_k &= r \sin \theta_1, \sin \theta_2, \dots, \sin \theta_{k-3} \sin \theta_{k-2} \sin \theta_{k-1}, \end{aligned}$$

where  $0 \leq \theta_1 \leq \theta_2 \leq \pi, \dots, 0 \leq \theta_{k-2} \leq \pi, 0 \leq \theta_{k-1} \leq 2\pi$  (see, for example Edwards, 1973). Values of  $\theta_1, \theta_2, \dots, \theta_{k-2}, \theta_{k-1}$  are randomly selected from independent uniform distribution such that  $\theta_i \sim U(0, \pi), i = 1, 2, \dots, k - 2; \theta_{k-1} \sim U(0, 2\pi)$ . For chosen  $r$ , we obtain  $x_1, x_2, \dots, x_k$  which are used to evaluate  $\eta(\mathbf{x})$ . A large number of points are chosen in this manner to obtain a sample,  $H(r)$  consisting of SPV values on  $S(r)$ . Then the quantiles of  $H(r)$  are obtained. Plots of the Quantiles of  $H(r)$  versus  $p$ , can be obtained for any value of  $r$  within the region  $R$ . The first quartile ( $p = 0.25$ ), the median ( $p = 0.50$ ), the third quartile ( $p = 0.75$ ) and so on can be obtained from the plots. Also, using this method the prediction variance properties of more than one design can be compared by superimposing their respective quantiles of  $H(r)$  for each value of  $r$  inside  $R$ . This is called the combined quantile plots. We shall use the combined quantile plots in comparing the different variations of the ccd.

## 2.3 Quantiles of the Variations of Partially Replicated ccd

For the preference of replicating the cube to star points of the ccd, we shall consider the following variations; see Nduka & Chigbu (2014)

one cube plus two replicated stars,  $n_\alpha = 2$

two replicated cubes plus one star,  $n_F = 2$

one cube plus three replicated stars,  $n_\alpha = 3$

three replicated cubes plus one star,  $n_F = 3$

one cube plus four replicated stars,  $n_\alpha = 4$

four replicated cubes plus one star,  $n_F = 4$

Consider using the ccd to fit ( 1 ). The fitted model is shown in ( 3 ). The quantiles of the SPV values of each variation are considered for  $k = 2, 3, 4$  and  $\alpha = 1$ . At  $\alpha = F^{\frac{1}{4}}$ , each variation is considered for  $k = 3$ . Also, for each  $k$ , 10 000 points are randomly generated from uniform distribution on  $S(r)$  by using  $\theta_i \sim U(0, \pi)$ , and  $\theta_{k-1} \sim U(0, 2\pi)$ . Then we obtain the plot of the quantiles of  $H(r)$  versus  $p$  for the aforementioned variations. For illustrations, values of  $r$  are selected for each  $k$  and the combined plots for the variations are shown in figures 1-4. We used MATLAB to generate uniform random numbers, obtain SPV's, the quantiles of  $H(r)$  and the combined quantile plots. MATLAB is powerful and efficient in handling matrices and graphics.

## 3. Results

### 3.1 Prediction Variance of the Partially Replicated ccd at $\alpha = 1$

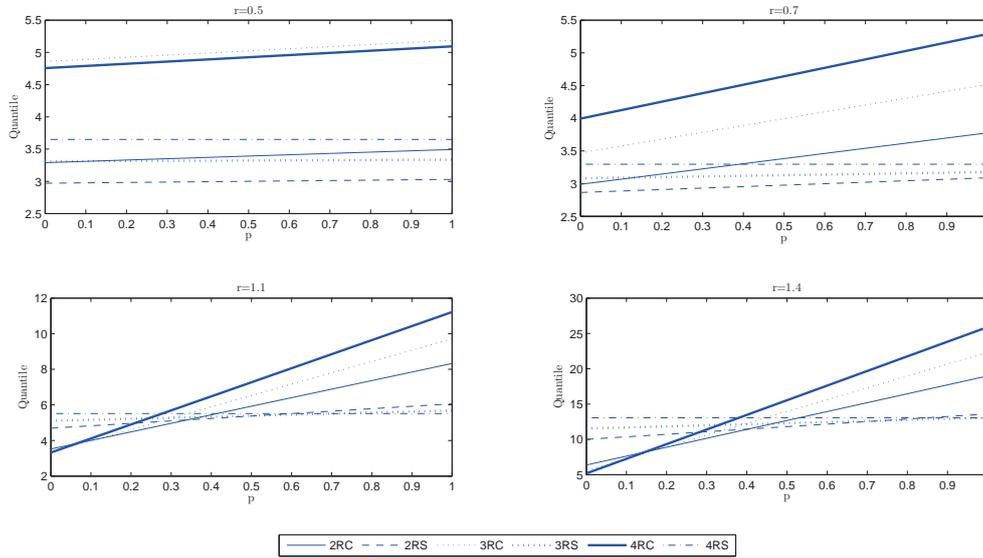


Figure 1. Combined quantile plots of replicated cubes plus one star and the replicated stars plus one cube for  $k = 2$ : RC implies replicated cube plus one star. RS implies one cube plus replicated star.

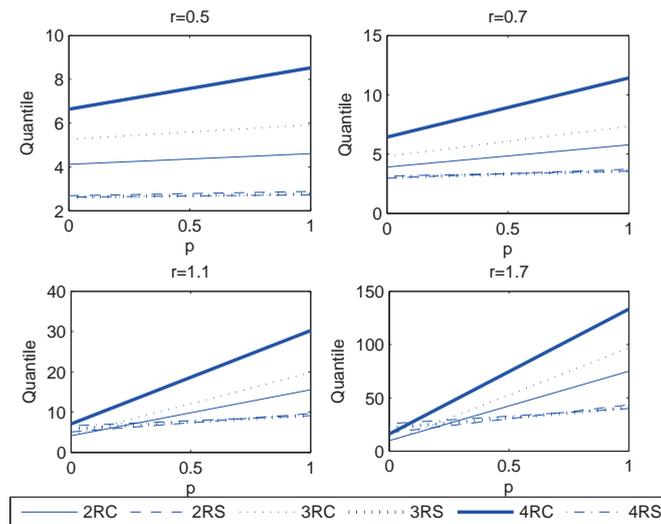


Figure 2. Combined quantile plots of replicated cube plus one star and the one cube plus replicated star for  $k = 3$

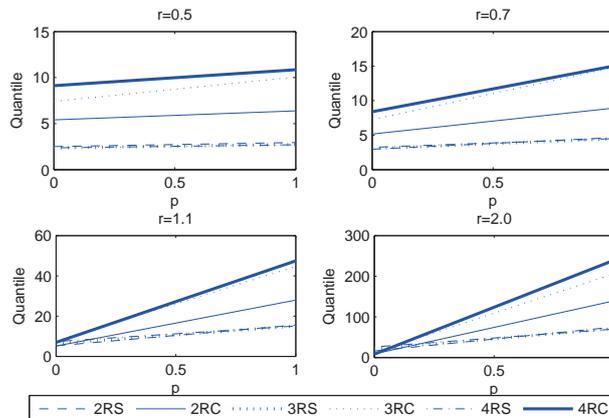


Figure 3. Combined quantile plots of replicated cube plus one star and the one cube plus replicated star for  $k = 4$

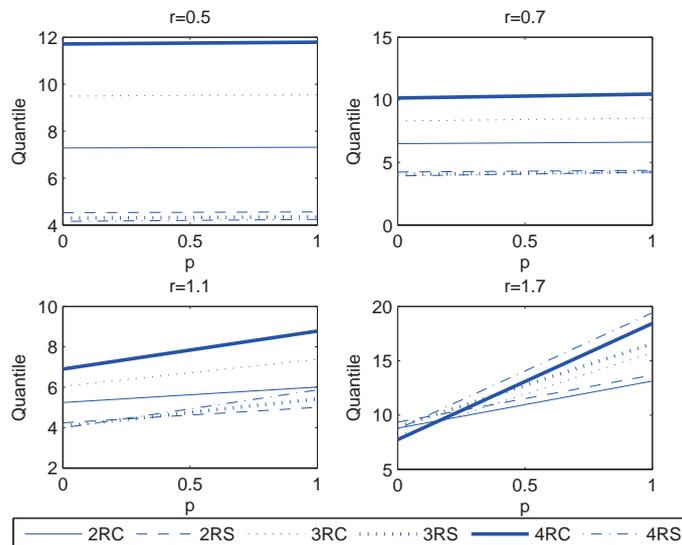


Figure 4. Combined quantile plots of replicated cube plus one star and the one cube plus replicated star for  $k = 3$  and  $\alpha = 1.682$

When  $k = 2$ , the combined quantile plots (figure 1) show that one cube plus two replicated star  $n_\alpha = 2$  has lower quantile values than the two replicated cubes plus one star  $n_F = 2$  for  $r = 0.5$  and  $0.7$ . For  $r = 1.1$  and  $1.4$ ,  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  have lower quantile values than  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  respectively for  $p \leq 0.4$ . But for  $p > 0.4$ ,  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  perform better than  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  respectively. Also, the plots of the replicated cubes plus one star depict large dispersion. For  $r = 0.5$  and  $r = 0.7$ , the quantiles of the SPV of  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  are lower than that of  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  respectively. When  $r = 1.1$  and  $1.7$ , the quantiles of the SPV of  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  are lower than that of  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  respectively for about  $p > 0.2$ , (see, figure 2). As seen in figure 1, figure 2 shows that the quantile plots of the replicated cubes plus one star depict dispersion. For  $k = 4$  (figure 3), the quantile values of  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  at  $r = 0.5, 0.7$  and  $1.1$ , are smaller than  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  respectively. When  $r = 2.0$ ,  $n_\alpha = 4$  performs better than  $n_F = 4$ , but the quantile values of  $n_F = 2$ ,  $n_F = 3$  are slightly lower than that of  $n_\alpha = 2$ ,  $n_\alpha = 3$  respectively for  $p < 0.1$ , but higher than the quantile values of  $n_F = 2$ , and  $n_F = 3$  for other values of  $p$ .

### 3.2 Prediction variance of the partially replicated rotatable ccd ( $\alpha = 1.682$ )

To illustrate, we considered a rotatable ccd with three variables of interest,  $k = 3$ . For  $r = 0.5, 0.7$  and  $1.1$ , (figure 4),  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  have smaller quantile values than  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$ . These variations of the partially replicated ccd show little dispersion. For  $r = 1.7$ , the quantile plots depict large variations and  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$  have lower quantile values than the  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$ .

## 4. Discussion

### 4.1 Comparison of the Prediction Variance of Replicated Cube Plus One Star with One Cube Plus Replicated Star ccd at $\alpha = 1$

Smaller values of the SPV are preferred to larger values. As seen in section 3, the one cube plus replicated star have smaller SPV than the replicated cube plus one star for  $r = 0.5$  and  $0.7$ , (figure 1-3). This implies that the one cube plus replicated star perform better than the replicated cube plus one star. For  $r = 1.1$  and  $1.4$ , ( $k = 2$ ), one cube plus replicated star perform better than the replicated cube plus one star as one moves towards the design perimeter. This is also applicable to  $r = 1.7$  and  $2.0$ . In addition to smaller SPV, the one cube plus replicated star show little or no dispersion. That is SPV is constant at all points which are equidistant from the centre of the design or near rotatable. (see also, Khuri, Kim & Um, 1996).

### 4.2 Comparison of the Prediction Variance of Replicated Cube Plus One Star with One Cube Plus Replicated Star of a Rotatable ccd.

For  $r = 0.5, 0.7$  and  $1.1$ , (figure 4)  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$  have lower quantile values than  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$ , hence the former is preferred to the latter. However, for  $r = 1.7$ ,  $n_F = 2$ ,  $n_F = 3$  and  $n_F = 4$ , perform better than  $n_\alpha = 2$ ,  $n_\alpha = 3$  and  $n_\alpha = 4$ , but the quantile plots depict large dispersion.

## 5. Conclusion

The prediction variance of the replicated cube plus one star and one cube plus replicated stars have been considered. The quantile plot is useful in design comparison because it shows the distribution of the different designs in the region of interest. It is not limited to only maxima and minima values. For the variables of interest considered here, the performance of the one cube plus replicated star variation in a face-centred ccd is preferred to the replicated cube plus one star variation. Even where the replicated cube plus one star perform better than the one cube plus replicated star, the latter still show a pronounced superiority as one moves towards the design perimeter. In rotatable ccd, where the replicated cube plus one star perform better than the one cube plus replicated star, the plots depict large dispersion of SPV. Therefore, one cube plus replicated star is preferred to replicated cube plus one star. The result obtained here is in consonance with the result obtained in Hader and Park (1978).

## References

- Atkinson, A. C., & Donev, A. N. (1992). *Optimum experimental designs*. New York: Oxford University Press.
- Borkowski, J. J. (1995). Spherical prediction-variance properties of central composite designs and Box-Behnken designs. *Technometrics*, 37(4), 399-410. <http://dx.doi.org/10.1080/00401706.1995.10484373>
- Chigbu, P. E., & Ohaegbulam, E. U. (2011). On the preference of replicating factorial runs to axial runs in restricted second-order designs. *Journal of Applied Sciences*, 11, 3732-3737. <http://dx.doi.org/10.3923/jas.2011.3732.3737>
- Dean, A., & Voss, D. (1999). *Design and analysis of experiments*. New York: Springer-verlag.
- Draper, N. R. (1982). Centre points in second-order response surface designs, *Technometrics*, 24(2), 127-133. <http://dx.doi.org/10.1080/00401706.1982.10487734>
- Edwards, C. H., Jr. (1973). *Advanced calculus of several variables*. New York: Academic Press.
- Giovannitti-Jensen, A., & Myers, R. H. (1989). Graphical assessment of the prediction capability of response surface designs, *Technometrics*, 31(2), 159-171. <http://dx.doi.org/10.1080/00401706.1989.10488510>
- Hader, R. J., & Park, S. H. (1978). Slope-rotatable central composite designs. *Technometrics*, 20, 413-417. <http://dx.doi.org/10.1080/00401706.1978.10489695>
- Khuri A. I., Kim, H. J., & Um, Y. (1996). Quantile plots of the prediction variance for response surface designs. *Computational Statistics and Data Analysis*, 22, 395-407.
- Li, J., Liang, L., Borrer, C. M., Anderson-Cook, C., & Montgomery, D. C. (2009). Graphical summaries to compare prediction variance performance for variations of the central composite design for 6 to 10 factors. *Quality Technology of Quantitative Management*, 6(4), 433-449.
- Lucas, J. M. (1976). Which response surface design is the best: A performance comparison of several types of quadratic response surface designs in symmetric regions. *Technometrics*, 18(4), 411-417. <http://dx.doi.org/10.1080/00401706.1976.10489472>
- Montgomery, D. C. (2013). *Design and analysis of experiments* (8th edition). Massachusetts: John Wiley.
- Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments* (3rd edition). New Jersey: John Wiley and Sons, Inc.
- Nduka, U. C., & Chigbu, P. E. (2014). On optimal choice of cube and star replications in restricted second-order designs. *Communications in Statistics-Theory and Methods*, 43, 4195-4214. <http://dx.doi.org/10.1080/03610926.2012.725499>
- Ukaegbu, E. C., & Chigbu, P. E. (2015). Comparison of the prediction capabilities of partially replicated central composite designs in cuboidal region. *Communications in Statistics-Theory and Methods*, 44, 406-427. <http://dx.doi.org/10.1080/03610926.2012.745561>

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).