Comparisons of Alternative Axial Distances for Cuboidal Regions of Central Composite Designs Using D and G Efficiencies

[1,2]Onyishi, Linus Ifeanyi [2]M.E. Nja [2]E.O. Effanga [3]F.C. Eze

[1]Department of Statistics, University of Calabar, Calabar, Nigeria

[2]Department of Mathematics and Statistics, Federal Polytechnic, Nasarawa, Nasarawa State, Nigeria

[3]Department of Statistics, Nnamdi-Azikiwe University, Awka, Nigeria

fc.eze@unizik.edu.ng

Abstract

In this study, three axial distances are proposed as alternatives to the existing axial distances of the Central Composite Design (CCD) in cuboidal design regions with the aim of providing formidable alternatives to the existing axial distances of the CCD whose prediction properties are less extreme and more stable in the cuboidal design regions. The three alternative axial distances, namely the arithmetic, harmonic and geometric axial distances for cuboidal regions, were developed algebraically based on the concepts of the three Pythagorean means. The strengths and weaknesses of the alternative axial distances were validated by comparing their performances with the existing axial distances in the cuboidal regions. The D- and G-efficiencies are used for comparison. The cuboidal region shows that the three alternative axial distances are consistently better in terms of the D- and G-efficiencies.

Key words: Axial Distances; Central Composite Design; D-Efficiency; G-Efficiency; Pythagorean Means.

1. Introduction

The Central composite design (CCD) emanated from the response surface designs and is the most popular and commonly used classes of experimental design for fitting a second-order response surface model given as

$$y_{ij} = \beta_0 + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j}\sum \beta_{ij} x_i x_j + \varepsilon_{ij} \quad (1)$$

where y_{ij} is the measured response; $x_i, i = 1,...,k$ are the input variables; $\beta_0, \beta_i, \beta_{ij}$ are the unknown parameters and ε_{ij} is the random error with mean zero and variance σ^2.

Generally, the CCD consists of a number of 2^k factorial (or fractional factorial of resolution V) points with r_f (number of replication at each factorial point), 2k axial or star runs and n_c (number of replicated center point). There are two parameters in the CCD design that must be specified: the distance α of the axial runs from the design center and number of center points n_0. The choice of α value specifies the type of central composite design and these leads to the various classes of central composite designs; Spherical central composite design (SCCD), Rotatable central composite design (RCCD), Orthogonal central composite design (OCCD) and Face centered cube design (FCCD). These designs are selected based on the choice of the axial, cube and center points with the extent of replication (Eze and Ngonadi 2018). Many second-order response surface designs exist of which the central composite design (CCD) is one of them. The central composite design was developed by Box and Wilson (1951) and has remained the most popular and practically useful class of second-order response surface designs. The symmetry and flexibility offered by the structure of the design give substantial advantage in prediction variance characterization and parameter estimation. The CCD exists for $k \geq 2$ in spherical and cuboidal regions, where k is the number of factors (independent variables of interest). According to Borkowski (1995), Li et al (2009) and Chigbu et al (2009), the structure of the CCD has three components: the factorial (cube) component which is at least a resolution V design, the star (axial) component at distance, α, from the centre of the design along each axis, and the centre point located at the centre of the design space. A resolution V design is a design in which two-factor interactions are aliased with three-factor interactions but no main effect or two-factor interaction is aliased with another main effect or two-factor interaction. That is, main effects and the two-factor interactions do not have other main effects and two-factor interactions as their aliases. Hence, for a resolution V design, the shortest word in the defining relation must have five letters.

The cube or factorial component of the CCD has $f = 2^{k-q}$ full (q = 0) or fractional (q >0) factorial number of runs, where q is an integer. The number of runs of the star component is *2k* which is augmented with n_0 centre points (the centre component). In other words, the CCD uses a total of $N = f + 2k + n_0$ number of runs to estimate the $p = (k+1)(k+2)/2$ number of model parameters. The cube (factorial) component has coordinates of the form, $(x_1, x_2, \ldots, x_k) = (\pm 1, \pm 1, \ldots, \pm 1)$; the star component has coordinates of the form, $(x_1, \ldots, x_k) = (\pm \alpha, 0, 0, \ldots, 0), \ldots, (0, 0, \ldots, \pm \alpha)$ while the centre is of the form, $(x_1, \ldots, x_k) = (0, 0, \ldots, 0)$.

The three components of the CCD play important but different roles in model parameter estimation. The at least resolution V full or fractional factorial component (the cube) contributes substantially to the estimation of the k linear terms and the $k(k-1)/2$ two-factor interaction terms of the second-order model. Only the factorial point contributes to the estimation of the interaction terms. The star component contributes to the estimation of the k quadratic terms of the second-order model. Without the star component, only the sum of the first-order terms can be estimated. The star component does not contribute to the estimation of the interaction terms. The centre component contributes to the estimation of pure error and estimation of quadratic terms (Wong, 1993).

Hence, the role of the star component of the CCD is very vital in response surface exploration using the central composite design. It is the introduction of the star component that gives the CCD the second-order design status since the factorial component must be a first-order design. Therefore, the location of the star points on the design region strongly influences the performance of the CCD. Usually, the star points are located at distance, α from the centre of the design region. This distance, α from the centre along the axes of the CCD is called the axial distance. Some axial distances exist which leads to the classification of the CCD according to the type of axial distance. For $\alpha = \sqrt{k}$, we have the spherical CCD, $\alpha = \sqrt[4]{f}$ gives the rotatable CCD, $\alpha = \sqrt[4]{k}$ gives the practical CCD and $\alpha = 1$ gives the Face-centered CCD.

The distinct structures of the central composite designs are characterized by the choice of the axial distance. According to Li *et al.* (2009), the axial distance defines the placement of the star points in response surface exploration using the CCD which consequently have substantial influence on the distribution of the design's prediction variance in the design region. Different axial distances have been proposed for different experimental purposes involving the CCD and each axial distance has specific effect on the structure and property of the CCD.

When the region of interest is cuboidal, Box and Wilson (1951) proposed that the axial distance should be $\alpha = 1$. This locates the star points at the centre of the faces of the cubes. Therefore, the CCD with this axial distance is popularly called the Face-centered central composite design or the Face-centered cube.

Liet et al (2009) identified that as the number of factors increases, alpha values could give impracticable axial distances. Anderson and Whitcomb (2005) proposed the practical alpha, $\alpha = \sqrt[4]{k}$ as a compromise for cuboidal alpha. This also offers reasonable variance inflation factor (VIF) as k increases (Li *et al.*, 2009).

Dykstra (1960) extended the rotatable axial distance to accommodate the replication of the cube and star components of the CCD. He gave the rotatable alpha as $\alpha = \sqrt[4]{\frac{n_1 f}{n_2}}$, where n_1 and n_2 are respectively, the number of replications of the cube and star components of the CCD. The rotatable alpha has been used extensively in the evaluation of the CCD for experimental purposes, Dykstra (1960), Draper (1982), Myers *et al.* (2009) and Ukaegbu and Chigbu (2015a). Though the rotatable axial distance may give desirable prediction variance properties, the level may be impractical for the factors of interest. For industrial experiments requiring large number of factors, say *k* = 20 factors, the star points will have rotatable axial distance of 4.471 (Li *et al.*, 2009). This value may not be feasible considering the fact that the remaining factors in the experiment are set at the levels of −1 and +1. This may degenerate further with the replication of the cube or both the cube and star components.

Ukaegbu (2017) has also shown that the CCD with practical axial distance did not compete favourably with the CCD with cuboidal α in a baking experiment requiring the choice of appropriate axial distance for the CCD to optimize the experiment. The practical axial distance is applicable in the spherical and cuboidal regions.

In this study, we will develop a set of alternative axial distances for the central composite designs in cuboidal design regions. The alternative sets of axial distances will be evaluated and compared with the existing axial distances in cuboidal design regions using D and G efficiencies. With the development of statistical software such as the Design Expert, which is based on numerous programs and algorithms, the computation of D and G efficiencies has been made easy.

2. Methodology

The axial distances are developed for the CCD. The axial distances are developed under the three classical Pythagorean means, the arithmetic mean, harmonic mean and geometric mean of the available axial distances in cuboidal regions. The arithmetic mean of a set of random variables, $\eta_1, \eta_2, \ldots, \eta_r$, is given by $\Lambda_a = r^{-1}\sum_{i=1}^{r}\eta_i$. The harmonic mean of the same set of random variables is $\Lambda_h = r\left[\sum_{i=1}^{r}\frac{1}{\eta_i}\right]^{-1}$ while their geometric mean is $\Lambda_G = \left[\prod_{i=1}^{s}\eta_i\right]^{\frac{1}{s}}$.

The axial distances developed from these functions are for *k* = 3 to 8 factors. These designs were evaluated using the *D*- and *G*-efficiency design evaluation criteria as the single-value criteria. The design that has the highest *D*- and *G*-efficiency values is the most efficient with respect to the particular efficiency criteria.

2.1 Prediction Variance

Characterization of the prediction variance is very important in comparing competing designs. A good design will have minimum prediction variance distributed throughout the entire design space. At a point, **x**, in the design space, the prediction variance is

$$\mathrm{var}[y(\mathbf{x})] = \sigma^2 \mathbf{x'^m}(\mathbf{X'X})^{-1}\mathbf{x^m}; \quad (2)$$

where, $(\boldsymbol{X'X})^{-1}$ is the inverse of the information matrix of a second-order response surface design whose design matrix is $\boldsymbol{X}$ and $\mathbf{x'^m} = (1, x_1, \ldots, x_k; x_1^2, \ldots, x_k^2; x_1x_2, \ldots, x_{k-1}x_k)$ is the vector of design points in the design space expanded to model form by classifying the coordinates of the design points into linear, quadratic and mixed (interaction) components of the model. By multiplying equation (2) by N gives the scaled prediction variance (SPV),

$$\frac{N\,\mathrm{var}[y(\mathbf{x})]}{\sigma^2} = N\mathbf{x}'^{\mathbf{m}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{\mathbf{m}}. \quad (3)$$

Traditionally, dividing equation (3) by σ^2 eliminates the unknown parameter which makes it impossible to use the equation to access the prediction variance characteristics of a second-order response surface design. However, in the literature, some authors arbitrarily assume $\sigma^2 = 1$ in order to eliminate it (Onukogu, 1997). In industrial settings, some experimenters prefer the standardized or unscaled prediction variance (UPV) for design evaluation. The unscaled prediction variance is given by

$$\frac{\mathrm{var}[y(\mathbf{x})]}{\sigma^2} = \mathbf{x}'^{\mathbf{m}}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}^{\mathbf{m}}. \quad (4)$$

The preference for UPV is because it provides the researcher the platform to assess the increase in precision (not influenced by the cost of experimentation, *N*) obtained from using larger designs by looking at the unscaled prediction variance. Usually, experimenters determine the cost of experimentation in response surface exploration through the total number of runs, *N*, of the experiment.

2.2 D-Efficiency

This is based on the determinant of $X'X$ which is inversely proportional to the square of the volume of the confidence region on the regression coefficients. It indicates how well the set of coefficients are estimated. According to Eze and Ngonadi (2018), a smaller $|X'X|$ or equivalently, a larger $|(X'X)^{-1}|$ implies poorer estimation of the regression coefficients in the model. The goal of D-optimality is to maximize $|X'X|$ or equivalently minimize $|(X'X)^{-1}|$ where X is the design matrix.

2.3 G-Efficiency

A G-optimality and the corresponding G-efficiency emphasize the use of designs for which the maximum $N\,\mathrm{var}[\hat{y}(x)]/\sigma^2$ in the region of the design is not too large. Hence G-optimality is based on

$v(x) = Nf'(x)(X'X)^{-1}f(x) = N[\hat{y}(x)]/\sigma^2$ Box and Hunter (1957) (5)

The aim is to minimize the maximum prediction variance in the design region.

Where X is the design matrix, x is any point in the design region, $f(x) = [f_1(x), \ldots, f_p(x)]'$ is a vector of p-real valued functions based on parameter model terms, and N is the design size.

Hence, Minimize $\max_{x \in X}\{Nf'(x)(X'X)^{-1}f(x)\}$

The G-efficiency is

$$\frac{100p}{N\hat{\sigma}^2_{\max}} \quad (6)$$

For the *D*- and *G*-efficiencies, the closer the values are to 1 (that is, 100%), the better the design. With the development of statistical softwares such as the JMP, Design Expert, and so on, which are based on numerous programs and algorithms, the computation of *D* and *G*-efficiencies has been made easy.

3. Results and Discussions

A set of axial distances were developed for the cuboidal design region for the evaluation of the central composite designs (CCD) in this section.

3.1 Axial distances for Cuboidal Region

The two existing axial distances applicable to the cuboidal region is the cuboidal alpha, $\alpha = 1$, which defines the face-centered CCD and the practical alpha, $\alpha = \sqrt[4]{k}$ (*k* is as earlier defined). Then the three Pythagorean means are obtained.

For the arithmetic alpha,

$$\alpha_{AC} = s^{-1}\left[1 + k^{\frac{1}{4}}\right], s = 2. \quad (7)$$

The harmonic alpha for the cuboidal region is

$$\alpha_{HC} = s\left[1 + \frac{1}{\sqrt[4]{k}}\right]^{-1}, s = 2. \quad (8)$$

The geometric alpha is

$$\alpha_{GC} = \left[1 \times \sqrt[4]{k}\right]^{1/s} = (k)^{1/4r}, s = 2. \quad (9)$$

Equations (7), (8) and (9) are the three alternative axial distances of the CCD in the cuboidal region. Catalogues of the values of these alternative axial distances with those of the existing axial distances, the cuboidal, α_C, spherical, α_S, practical, α_P, and rotatable, α_R, axial distances, from which they were derived are presented in Tables 1 for cuboidal regions.

Table 1: Catalogue of Alpha Value for the Cuboidal Region

k	**f**	α_c	α_P	α_{AS}	α_{HS}	α_{GS}
2	2^2	1.0000	1.1892	1.0946	1.0864	1.0905
3	2^3	1.0000	1.3161	1.1580	1.1365	1.1472
4	2^4	1.0000	1.4142	1.2071	1.1716	1.1892
5	2^5	1.0000	1.4954	1.2477	1.1985	1.2229
6	2^6	1.0000	1.5651	1.2826	1.2203	1.2511
7	2^{6-1}	1.0000	1.5651	1.2826	1.2203	1.2511
8	2^7	1.0000	1.6266	1.3133	1.2386	1.2754
9	2^{7-1}	1.0000	1.6266	1.3133	1.2386	1.2754
10	2^{7-2}	1.0000	1.6266	1.3133	1.2386	1.2754
	2^8	1.0000	1.6818	1.3409	1.2542	1.2968

	2^{8-1}	1.0000	1.6818	1.3409	1.2542	1.2968
	2^{8-2}	1.0000	1.6818	1.3409	1.2542	1.2968
	2^{9}	1.0000	1.7321	1.3661	1.2680	1.3161
	2^{9-1}	1.0000	1.7321	1.3661	1.2680	1.3161
	2^{9-2}	1.0000	1.7321	1.3661	1.2680	1.3161
	2^{9-3}	1.0000	1.7321	1.3661	1.2680	1.3161
	2^{10}	1.0000	1.7783	1.3892	1.2801	1.3335
	2^{10-1}	1.0000	1.7783	1.3892	1.2801	1.3335
	2^{10-2}	1.0000	1.7783	1.3892	1.2801	1.3335
	2^{10-3}	1.0000	1.7783	1.3892	1.2801	1.3335

3.2 Design Efficiencies for Comparison

The D- and G-efficiency values are computed for five variations of the CCD in cuboidal region. The variations of the CCD are based on the various axial distances defining the central composite designs under comparison. The D- and G-efficiencies are used to measure any improvement on the performances of the designs by the alternative axial distances. The closer the efficiency value is to 100 percent, the more preferable the design. The designs having the same number of experimental runs makes it easier for comparison as no design will have undue advantage over the others with respect to the number of runs. The number of factors under consideration here is k = 2 to 8 factors. Each set of factors was evaluated with three centre points in line with the recommendations of Montgomery (2013) which stated that for CCD, 3-5 centre points area acceptable. The values of the D- and G-efficiencies for the axial distances for the cuboidal region are presented in Tables 2 and 3, respectively.

Table 2: D-efficiency Values with Three Centre Points in Cuboidal Region

k	***F***	***N***	α_C	α_P	α_{AC}	α_{HC}	α_{GC}
2	2^2	11	42.84	44.77	46.43	46.03	46.36
3	2^3	17	41.30	42.95	46.83	46.08	46.45
4	2^4	27	42.10	44.07	49.05	47.88	48.46
5	2^5	45	43.30	48.16	51.24	49.71	50.47
6	2^{6-1}	47	43.20	49.77	51.66	49.84	50.74
7	2^{7-1}	81	45.11	51.28	54.00	51.95	52.97

8	2^{8-2}	83	45.88 51.40 55.07 52.82 53.93

Table 3: G-efficiency Values with Three Centre Points in Cuboidal Region

k	*F*	*N*	α_C	α_P	α_{AC}	α_{HC}	α_{GC}
2	2^2	11	68.71	57.65	72.18	71.79	72.03
3	2^3	17	74.00	69.42	76.37	76.02	76.20
4	2^4	27	84.30	73.91	85.86	85.57	85.71
5	2^5	45	95.01	82.20	92.22	92.52	92.37
6	2^{6-1}	47	88.14	84.14	88.95	88.74	88.85
7	2^{7-1}	81	89.20	86.30	87.82	88.12	87.97
8	2^{8-2}	83	94.37	91.96	94.88	94.75	94.81

3.3 G-efficiency Values with Three Centre Points in Cuboidal Region

As depicted by the results in Table 2, the CCDs with arithmetic, harmonic and geometric axial distances are consistently the best with the highest D-efficiency values for all the factors considered in the cuboidal design region. Among the three alternative axial distances, the CCD with the arithmetic axial distance gives the highest D-efficiency values which are followed by the CCD with geometric axial distance and then, the CCD with harmonic axial distance giving the lowest percentage D-efficiency values. Also, the CCDs with arithmetic, harmonic and geometric axial distances provide the highest G-efficiency values as compared to the values of those of the cuboidal and practical axial distances. The only exceptions are for k = 5 and 7 where the CCD with cuboidal axial distance, α_C have the highest G-efficiency values. The CCD with arithmetic axial distance again provided the highest G-efficiency values for k = 3, 4, 6 and 8 factors among the three alternative axial distance. This is followed by the CCD with geometric axial distances and then, the CCD with harmonic axial distance. Obviously, the central composite designs with arithmetic axial distance would be the experimenter's choice should the D-efficiency criterion for response surface exploration. The same is also true of the CCD with arithmetic axial distance if the G-efficiency is the experimenter's choice for response surface exploration in cuboidal region. Following the results of the D- and G-efficiencies, the central composite designs with geometric and harmonic axial distance provide formidable alternatives to the CCD with arithmetic axial distances since in most cases, the results of the three CCDs only slightly differ.

4. Conclusion

The analytical forms of the arithmetic, harmonic and geometric means were extensively exploited in the development of the arithmetic, harmonic and geometric alphas as formidable alternative axial distances for the central composite design. The statistical properties of these alternative axial distances were explored using the D- and G-efficiencies. D-efficiency values of the arithmetic, harmonic and geometric axial distances are the highest among the five variations of the CCD evaluated for k = 2 to 8 factors. Also, the CCD with the three alternative axial distances have the highest G-efficiency values in the cuboidal region except for k = 5 and 8 where the CCD with cuboidal axial distance has the highest G-efficiency values.

In general and from the foregoing, the alternative axial distances, the arithmetic, harmonic and geometric axial distances, have yielded desirable results for the exploration of the response surfaces using the central composite design. The alternative axial distances are viable and formidable alternatives to the other existing axial distances for the central composite design in the cuboidal design region.

References

1. Anderson-Cook, C.M, Borr, C.M. and Montgomery, D.C. (2009), Response Surface Design Evaluation and Comparison, *Journal of Statistical Planning and Inference,* Vol. 139, 629-641. Doi: 10.1016/j.jsp2008.04.004.

2. Borkowski, J. J. (1995), Spherical Prediction Properties of Central Composite and Box Behnken Designs, *Technometrics,*Vol. 37 (4), 399 – 410.

3. Box, G. E. P. and Hunter, J. S., 1957. "Multifactor experimental designs for exploring response surfaces." *Annals of Mathematical Statistics,* vol. 28, pp. 195-241.

4. Box, G. E. P & Wilson, K. B. (1951), On the Experimental Attainment of Optimum Conditions, *Journal of the Royal Statistical Society,* Series B, Vol. 13, 1 – 45.

5. Chigbu, P. E., Ukaegbu, E. C. & Nwanya, J. C. (2009). On Comparing the Prediction Variances of Some Central Composite Designs in Spherical Region: A Review, *Statistica,* anno LXIX, n.4, 285-298.

6. Draper, N. R. (1982), Centre Points in Second-Order Response Surface Designs, *Technometrics,* Vol. 24 (2), 127 – 133.

7. Dykstra, O. Jr. (1960), Partial Duplication of Response Surface Designs, *Technometrics*, Vol. 2 (2), 185 – 195.

8. Eze, F.C. and Ngonadi, L.O (2018). "Alphabetic Optimality Criteria for 2^k Central Composite Design." Academic Journal of Applied Mathematical Sciences vol.4, issue 9. pp 107-118.

9. Li, J., Liang, L., Borror, C. M., Anderson-Cook, C. M. & 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 and Quantitative Management,* Vol. 6 (4), 433 – 449.

10. Montgomery, D. C. (2013), *Design and Analysis of Experiments,* 8th Ed., John Wiley and Sons Inc. N.Y.

11. Onukogu, I. B. (1997), Foundations of Optimum Exploration of Response Surfaces, Ephrata Press, Nsukka.

12. Ukaegbu, E. C. (2017). Evaluation of Prediction Capabilities of Partially Replicated Variations of Some Central Composite Design, A PhD Thesis Submitted to the Department of Statistics, University of Nigeria, Nsukka.

13. Wong, W. K. (1993), Comparing Robust Properties of A, D, E and G-optimal Designs, *Computational Statistics and Data Analysis*, Vol. 18, pp. 441-448.