A GRAPHICAL STUDY ON THE MISSING DATA OF CENTRAL COMPOSITE DESIGN WITHIN A SPHERICAL REGION

A.R. Gokul[1] and M. Pachamuthu[2*]

•

[1]Research Scholar, Department of Statistics, Periyar University, Salem-636011, India
gokulmalar6@gmail.com
[2*]Assistant Professor, Department of Statistics, Periyar University, Salem-636011, India
pachamuthu@periyaruniversity.ac.in

Abstract

Robust missing observations have emerged as a crucial study area in statistical research. Response Surface Methodology (RSM), a recognized and extensively utilized area in experimental design, has determined that the absence of observations in an experiment can introduce complexity and interfere with the estimation of parameters. Previous literature reviews reveal that most studies on missing Central Composite Design (CCD) data were conducted using optimality and minimax loss criteria. Our study explores the spherical region of interest in the missing observation of CCD, represented through Variance Dispersion Graph (VDG) and Fraction of Design Space (FDS) graphs. Practitioners primarily focus on the region of interest rather than employing various alpha values. We investigate the predictive capabilities of each factorial, axial, and center missing design point against different radii(r) and fractions of the design space region, and we also measure relative G- and D- efficiency. We scrutinize various factors (k) from two to seven, including five center runs. Our research explores the region of interest in operating the experiment under robust conditions through visual aids of VDG and FDS graphs.

Keywords: Central Composite Design, Fraction of Design Space, Scaled Prediction Variance, Optimality, Variance Dispersion Graph

I. Introduction

Response Surface Methodology is a potent combination of statistical and mathematical techniques used for model building. It's specifically designed to evaluate the effects of several independent variables and determine their optimal values to achieve the most favorable results. This methodology is particularly beneficial when the goal is to optimize a product or process. It allows for a comprehensive understanding of the relationships between different variables and the response, facilitating efficient and effective optimization. The empirical model is constructed using data gathered directly from the system or process under study. *RSM* involves building empirical models using multiple linear regression and statistical techniques [12] In the literature review, various authors have conducted studies on the missing data of second-order composite designs. The details of these studies will be briefly discussed sequentially. Draper [8] examined the studies on sturdy methods for handling missing observations in response surface design and acknowledged the pioneer who formulated the parameter estimation equation. Indeed, Akhtar et

al. [2] introduced a minimax loss criterion to assess missing observations. This approach has become the most used method in response surface designs. It's a significant contribution to the field as it provides a robust and effective way to handle missing data, thereby improving the accuracy and reliability of the designs. Alrweili et al. [4] discussed the minimax loss criterion to develop more resilient models for missing data. They achieve this by integrating the most recent *CCD*s from *GSA* and *AEK*, both novel designs. Hayat et al. [10] delved into designs derived from regular and irregular structure subsets. They evaluate how these designs handle missing points using the minimax loss criterion. Furthermore, they scrutinize their alphabetic optimality and predictive capabilities through FDS plots depicting the variance in response difference. Alanazi et al. [3] introduced closed-form expressions that account for two missing observations. These expressions are based on α, the axial value utilized in *CCDs* that handle up to 10 factors. Hemavathi et al.[11] examined the ability of sequential third-order rotatable design to manage missing data without significant information loss. Additionally, the study quantifies the loss of information from one or two absent experimental runs at varying distances from the design's center. Park et al. [14] compared *CCD*, *SCD*, and *MinResV* designs. They focus on spherical regions with k = 3 to 7 factors, using the optimality criteria and the variance dispersion graph as benchmarks. Interestingly, their findings reveal that none of these designs consistently outperforms others. Li et al. [13] evaluated various *CCD*, *SCD*, and *MinResV* designs. These designs are applied to spherical and cuboidal regions with different axial values. To analyze the prediction variance properties of these designs, they use *FDS* plots and box plots. Onwuamaeze [15] employed graphical techniques like *VDG* and *FDS* plots. They use these methods to assess the prediction variance performance of *CCD*, *SCD*, and *MinResV* designs within the hypercube region. Ahmad et al. [1] compared Augmented Pairs (*AP*) designs and Subset designs based on standard optimality criteria and graphical criteria in spherical and cuboidal regions of experimentation, which provide more insight into the prediction performance of the designs. The article [17] employed *VDG* and *FDS* plots to depict the scaled prediction variance attributes of the second-order design and *G-* and *I-* optimality designs within a cuboidal area of interest. The study by *G. G.* Vining et al. [18]) involved a graphical method that plots the maximum and average mean squared prediction error across spheres of different radii within the design space.

This study thoroughly investigates the impact of missing data on the central composite design within a spherical region of interest. Utilizing *VDG* and *FDS* determines the robustness of the design in various regions based on missing factorial, axial, and center design points. The paper is segmented into various parts. The methodology is detailed in Section 2. Section 3 presents the findings and discussions, including the interpretation of *VDG* and *FDS* plots using spherical and rotatable alpha values and Section 4 concludes the study.

II. Methodology

I. Description of Second-Order Central Composite Design

In many cases where we apply RSM, the relationship between the predictor variables and the response might be unclear. A first-order model, while applicable in some cases, might not be able to accurately capture the curvature of the response function due to its linear nature. This is where higher-degree polynomial models, such as second-order models, come into play. These models can capture more complex relationships and better evaluate curvature in optimization experiments. They provide a more nuanced understanding of the data, allowing for more accurate predictions and more effective optimization.

For k quantitative factors denoted by $x_1, x_2, \ldots, x_k$, a second-order model is

$$y = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_{ii} x_i^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij} x_i x_j + \varepsilon \quad (1)$$

In this context, $\beta_0, \beta_i, \beta_{ii}$, and β_{ij}represent the intercept, linear, quadratic, and bilinear coefficients, respectively. ϵ_i is a random error term with a mean of zero, a variance of σ^2, and is independent for each pair of runs. The total number of parameters that need to be estimated, denoted as p, is calculated as$p = k + k + \binom{k}{2} + 1$. To ensure enough degrees of freedom to estimate these model coefficients, the number of runs, represented as n, must equal or exceed p.

CCD is extensively utilized to estimate second-order response surfaces. Since its introduction by [7], the *CCD* has been the subject of numerous studies and has seen widespread use in various fields. The flexibility in utilizing the *CCD* lies in choosing alpha (α) , which represents the axial distance and n_c, denoting the number of centre runs. The selection of these two parameters can be crucial. The operational region and the area of interest largely influence the determination of α. Here, we use spherical and rotatable α values, where the spherical value of α is set to the square root of the number of factors, represented as $\alpha_S = k^{1/2}$. Indeed, in αs, all the design points are situated on a common geometric sphere, and it is nearly rotatable, and Box et al. [6] suggested a value, $\alpha_R = F^{1/4}$, called rotatable α, where F is the number of factorial runs in a design. Indeed, a design is rotatable if the variance remains the same for all points that are an equal distance from the design's center.

II. Scaled Prediction Variance and Relative G- and D- Efficiency

Borkoski [5] have developed an analytical form for calculating scaled prediction variance values of *CCD* and Box-Behnken design (*BBD*). Spherical Prediction Variance (*SPV*) enables accurate prediction of response variables at different points of interest within the experimental area. The prediction variance at a point *x* is given by

$$v(x) = \frac{n \cdot var[\hat{y}_{(x)}]}{\sigma^2} = n \cdot X^{(m)'}(X'X)^{-1}X^{(m)} \quad (2)$$

The vector $X^{(h)}$ represents the array of coordinates of a point in the design space that has been magnified to align with the model form, where n is the quantity of experimental runs design, and σ^2 is the observation error.

G- optimality is a standard that seeks to reduce the highest possible variance in any forecasted value across the entire experimental domain. This efficiency can be understood as the proportion between the determinant of the information matrix for a specific design and the determinant of the information matrix for the best optimal design.

$$G_{\text{eff}} = \frac{p}{n \cdot MAX_{X \in R} v(x)} \quad (3)$$

In this context, p represents the estimated model's parameters, and n denotes the number of observations in the corresponding design. The term $MAX_{X \in R} v_{(x)}$signifies the maximum value of the variance of the predicted response. As a result, the relative *G*-efficiency, denoted as RE_G, is determined by the ratio of the $G_{eff}$for the reduced design to the G_{eff} for the complete design.

$$RE_G = \frac{G_{\text{eff (reduced)}}}{G_{\text{eff}}} = \frac{n \cdot MAX_{X \in R} v(x)}{n_r \cdot MAX_{X \in R} v(x)_{\text{reduced}}} \quad (4)$$

In this scenario, n represents the size of the complete design, while n_r denotes the size of the reduced design. Based on the definition of RE_G, a design that yields a higher value of RE_G would be more desirable. This is because a higher RE_G value indicates a more efficient design relative to the complete design. By utilizing equations (3) and (4), we are able to compute the relative G-efficiency value. These values are then presented in tables 1 and 2.

D efficiency is defined as maximizing the determinant of the information matrix minimizing the determinant of the inverse of the information matrix. Thus, relative D- efficiency is given as

$$RE_D = \left(\frac{|X'X|_{\text{reduced}}}{|X'X|} \right)^{\frac{1}{p}} \quad (5)$$

Where, p is the number of parameters of the model to be estimated, $|X'X|_{\text{reduced}}$ is the determinant of the information matrix of reduced design and $|X'X|$ is the determinant of the complete design matrix. A value approaching one will represent a minor loss, whereas a value below one will represent a more significant loss in model estimation. Through the application of equation (5), we are able to determine the relative D-efficiency value. These computed values are then listed in tables 1 and 2.

III. Variance Dispersion Graph and Fraction of Design Space

Giovannetti-Jensen and Myers [9] presented the concept of variance dispersion graphs to evaluate the comprehensive predictive capability of a *RSM* within a region of interest. A variance dispersion graph allows one to visualize the uniformity of the scaled variance of a predicted value in multidimensional space. It consists of three curves: the maximum, the minimum and the average scaled variance of a predicted value on a hypersphere. However, these *VDG* plots do not consider the fraction of the total design space between concentric spheres of radius *r* at different distances from the center of the design space. Variance Dispersion Graph handle the *SPV* on a sphere with radius *r*, but they overlook the volume related to this information. To gain insight into the complete picture of the prediction performance of a design, one should consider the volume. An *FDS* plot [16] is generated by taking a substantial number of samples, represented as n, from the entire design area and calculating the associated *SPV* values. The underlying concept is that the design quality improves if a greater portion of the design space is near the minimum *SPV* value.

Furthermore, a flatter line indicates a more stable design. The *FDS* plot effectively encapsulates the range and distribution of *SPV* values in the design space, facilitating the comparison of designs through a single curve. Furthermore, it [19] equips the investigator with a unique graph for contrasting designs or examining the characteristics of a particular design.

III. Result and Discussion

I. Relative *D*- and *G*- efficiency of *CCD k* = 2 to 7 using spherical and rotatable α

When a data point is missing in a Central Composite Design (CCD) with factor k = 5 and a factorial (1,1,-1,1,-1) run, the relative G efficiency is poorest. G-efficiency measures the precision of parameter estimates in a design. Specifically, relative G efficiency compares the efficiency of a design with missing observations to the efficiency of the same design without any missing data. When a data point is missing, the relative G efficiency decreases, indicating reduced precision in estimating model parameters.

Additionally, relative D efficiency approaches 1 in this scenario, indicating maximum determinant value. D-efficiency is related to the determinant of the information matrix (also known as the Fisher information matrix). Relative D efficiency compares the determinant of the information matrix for a design with missing observations to the determinant for the complete design. A value closer to 1 indicates better efficiency in terms of information content.

Axial missing points outperform factorial points in terms of both relative G and D efficiency. Interestingly, missing a center run has minimal impact compared to the no-missing scenario, except for relative G efficiency when k = 2. Overall, relative G and D efficiency provide valuable insights for comparing missing and complete designs

Table 1: *Relative D and G efficiency of CCD of factors k = 2 to 7 using spherical α*

Factors	Number of runs	Types of Missing runs	Alpha value (α_S)	Relative G efficiency (RE_G)	Relative D efficiency (RE_D)
K=2	13	None	1.41421	1.0000	1.0000
	12	Factorial (-1,1)		0.4062	0.8492
		Axial (0,1.41421)		0.4062	0.8492
		Centre (0,0)		0.3008	0.9635
K=3	19	None	1.73205	1.0000	1.0000
	18	Factorial (-1, -1,1)		0.3581	0.8975
		Axial (0,1.73205,0)		0.4292	0.9080
		Centre (0,0,0,)		1.0555	0.9779
K = 4	29	None	2	1.0000	1.0000
	28	Factorial (-1,1,1,1)		0.4316	0.9433
		Axial (0,-2,0,0)		0.4316	0.9433
		Centre (0,0,0,0)		1.0358	0.9852
K = 5	31	None	2.23607	1.0000	1.0000
	30	Factorial (1,1,-1,1,-1)		0.1491	0.9120
		Axial (0,0,0,2.23607,0)		0.5177	0.9537
		Centre (0,0,0,0,0)		1.0334	0.9894
K = 6	49	None	2.4495	1.0000	1.0000
	48	Factorial (-1,-1,-1,1,1,-1)		0.3800	0.9653
		Axial (0,0,0,2.44949,0,0)		0.4723	0.9699
		Centre (0,0,0,0,0,0)		1.0209	0.9921
K = 7	83	None	2.64575	1.0000	1.0000
	82	Factorial (1,-1,-1,-1,1,1,-1)		0.7106	0.9845
		Axial (0,0,0,0,2.64575,0,0)		0.4747	0.9792
		Centre (0,0,0,0,0,0,0)		1.0122	0.9938

Table 2 : *Relative D and G efficiency of CCD of factors k =3,4,5 and 7 using rotatable* α

Factors	Number of runs	Types of Missing runs	Alpha value (α_S)	Relative G efficiency (RE_G)	Relative D efficiency (RE_D)
K = 3	19	None	1.68179	1.0000	1.0000
	18	Factorial (-1,-1,1)		0.3486	0.8951
		Axial (0,1.68179,0)		0.4178	0.9108
		Centre (0,0,0,)		1.0555	0.9780
K = 5	31	None	2	1.0000	1.0000
	30	Factorial (1,1,-1,1,-1)		0.1251	0.9044
		Axial (0,0,0,2,0)		0.4722	0.9574
		Centre (0,0,0,0,0)		1.0331	0.9901
K = 6	49	None	2.37841	1.0000	1.0000
	48	Factorial (-1,-1,-1,1,1,-1)		0.3770	0.9651
		Axial (0,0,0,2.37841,0,0)		0.4888	0.9704
		Centre (0,0,0,0,0,0)		1.0209	0.9921
K = 7	83	None	2.82843	1.0000	1.0000
	82	Factorial (1,-1,-1,-1,1,1,-1)		0.5790	0.9457
		Axial(0,0,0,0, 2.82843,0,0)		0.5340	0.9239
		Centre (0,0,0,0,0,0,0)		1.0121	0.9790

II. Interpretation of VDG of factors k = 2 to 7 of spherical $\sqrt{k}$ alpha value

Figure 1 depicts all the factors of *SPV* distribution against radius; when k = 2, the trajectory of the maximum, average, and minimum *SPV* for both factorial and axial design points with missing observations is identical. This trajectory remains consistent for each design point, beginning at a radius of 0 and extending to $\sqrt{2}$, resulting in a *maxSPV* value of 24.13. In this case, both the original and missing center design points exhibit the same maximum, minimum, and average prediction variance across all radii.

When k = 3, excluding *Fmax* and *Amax*, all other design points for maximum, average, and minimum prediction variance display a certain level of consistency in their *SPV* values below 5 up to a radius $r \leq 1.20$. Past a radius of about 1.20, the curve steadily rises until the radius hits $\sqrt{3}$. *Fmax* can achieve an *SPV* as elevated as 39.40, while *Amax* can attain an *SPV* as high as 33.40. For factor k = 4, a center run is missing at the radius origin, exhibiting an *SPV* value of 7.03. As the radius extends to r = 1, this *SPV* value decreases to 5.87. Beyond r = 1, the *SPV* value escalates with the radius, reaching a *maxSPV* of 16.26. A factorial and an axial design point are absent, with a maxSPV of 37.98 at a radius $\sqrt{4}$. The pattern followed by factor k = 4 resembles factor k = 2.

For factor k = 5, the lack of a factorial observation leads to a significant *SPV* value of 190.2 at the radius $\sqrt{5}$. This suggests that insufficient data under these conditions leads to subpar predictive performance in the experiment. The axial design point's maximum prediction variance is 52.39 at the radius $\sqrt{5}$. The *None-max, None-avg, Cmax, Cavg*, and *Aavg* of all these prediction variance design points are almost identical for all radii with an *SPV* of ~ 26.15 or less, except for *Favg SPV*, which slightly deviates from another curve from a radius of 1.30. Despite misinformation, *minSPVs* follow the same path for all design points.

When k = 6, the absence of a factorial and an axial observation result in the maximum *SPV* staying below 10 from a radius of $0 \leq r \leq 1.20$. However, as the radius increases beyond 1.20, the *SPV* value rises, reaching a maximum *SPV* of 73.90 for factorial and 63.20 for axial at radius $\sqrt{6}$.

Specifically, we can observe that the *SPV* design points *Favg*, *Fmin*, *Aavg*, and *Amin* follow the same trajectory from a radius of $0 \leq r \leq 1.40$, with an *SPV* of less than 8.65. Beyond a radius of 1.40, the *minSPVs* diverge from the path, reaching a *minSPV* of 29.40. A similar pattern is observed for *None-max*, *avg*, *min* and *Cmax*, *avg*, *min*, but they start from different *SPV* values at radius 0 and travel closely up to a radius of approximately 1.40. Beyond a radius of 1.40, the *minSPVs* diverge up to a radius of $\sqrt{6}$.

For factor $k = 7$, the lack of an axial design point leads to a predictive performance that is inferior to the factorial. The axial design point has an *SPV* of 71.85, while the factorial design point has an *SPV* of 53.75 at the radius $\sqrt{7}$. Excluding the center design point *SPV*, all other points start with an *SPV* of 16.20 at a radius of 0 and maintain the same path up to a radius slightly less than 1.0. Beyond a radius of 1.0, the prediction variances of each design point diverge from each other. It can be observed that the average and minimum *SPV* of various design points falls within an *SPV* interval of 35.31 to 33.50 at radius $\sqrt{7}$.

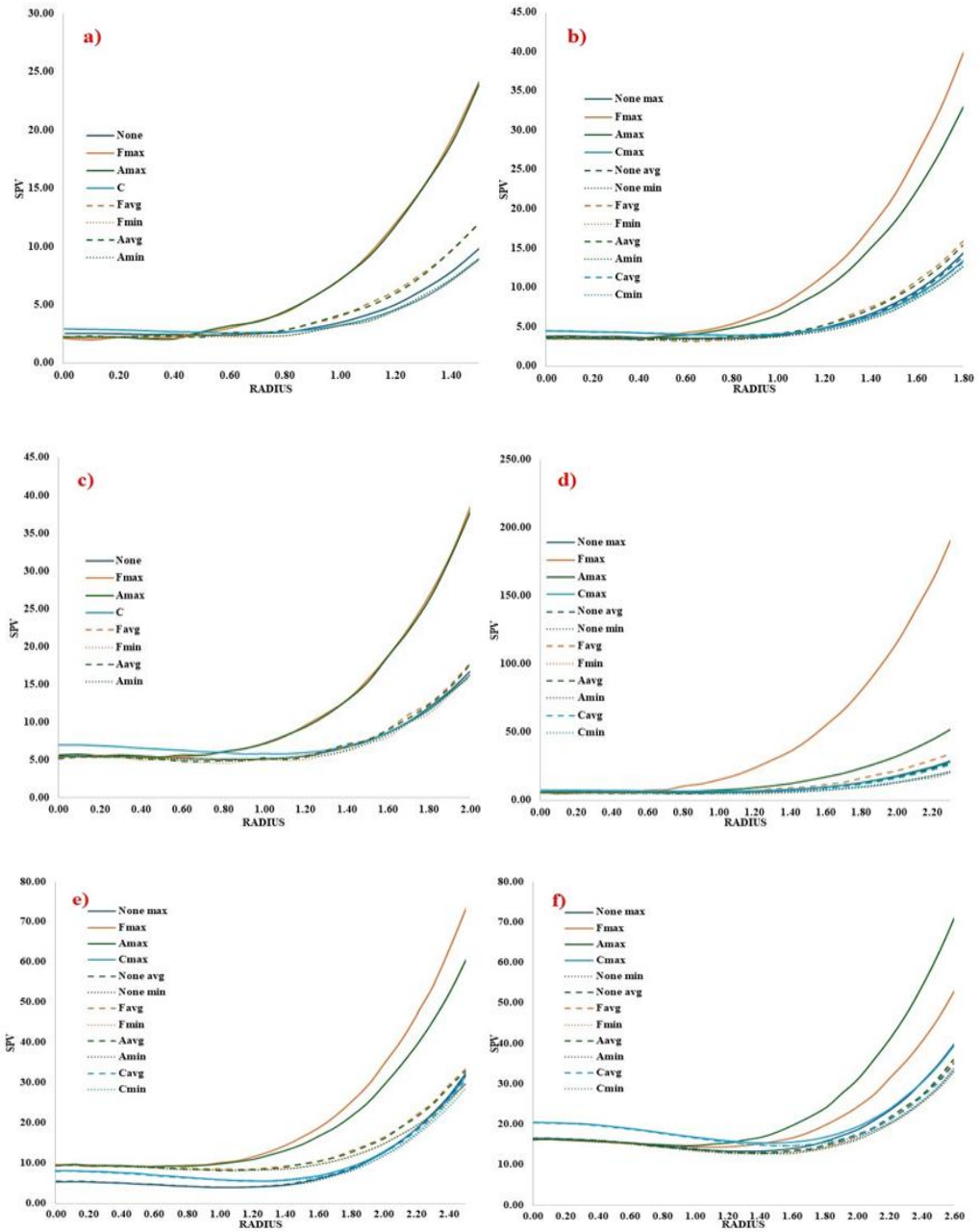

Figure 1: *(a) VDG for CCD K = 2. (b) VDG for CCD K = 3. (c) VDG for CCD K = 4. (d) VDG for CCD K = 5. (e) VDG for CCD K = 6. (F) VDG for CCD K = 7*

III. Interpretation of VDG of factors K = 3, 5, 6 and 7 of rotatable alpha value

Figure 2 depicts all the factors of *SPV* distribution against radius; For the factor k = 3, the absence of a center run results in the lowest *maxSPV* of 11.18. Excluding *Fmax* and *Amax*, which have SPVs of 34.11 and 28.76, respectively, all other average and minimum SPVs and design points with no missing data maintain an *SPV* value of less than 3.63 up to a radius of less than 1.00. Beyond a radius of 1.00, these design points begin to diverge slightly from each other.

For factor k = 5, when a factorial observation is absent up to a radius of 0.7, the SPV stays below 8.43. However, once the radius surpasses 0.7, the SPV undergoes a substantial increase, reaching a peak of 199.90 at $\sqrt{5}$. The lack of an axial observation results in a *maxSPV* of 8.41 at a radius of 1.00, and as the radius grows, the *SPV* steadily increases, attaining a *maxSPV* of 55.23 at $\sqrt{5}$. Apart from the two scenarios mentioned earlier, all other situations, such as no missing design points, a missing Centre run, average, and *minSPVs*, lie within the range of 33.71 to 24.81 at a radius of 2.23.

For factor k = 6, the lack of a factorial and an axial design point in an experimental setup results in nearly identical *SPV* values up to a radius of less than 1.20. Beyond this point, the *SPV* value slowly diverges, reaching 74.79 for factorial and 65.00 for axial at a radius of $\sqrt{6}$. For the factor k = 7, the absence of an axial observation leads to a lower prediction performance than when a factorial observation is missing. However, for radii less than 1.60, both design points maintain a nearly identical SPV value of 17.27. Factors k = 6 and 7 display the same conditions for no missing observation, average, and *minSPVs*, following the same path across all radii. The absence of a center run does not exhibit a consistent increasing trend across all radii; instead, it varies.

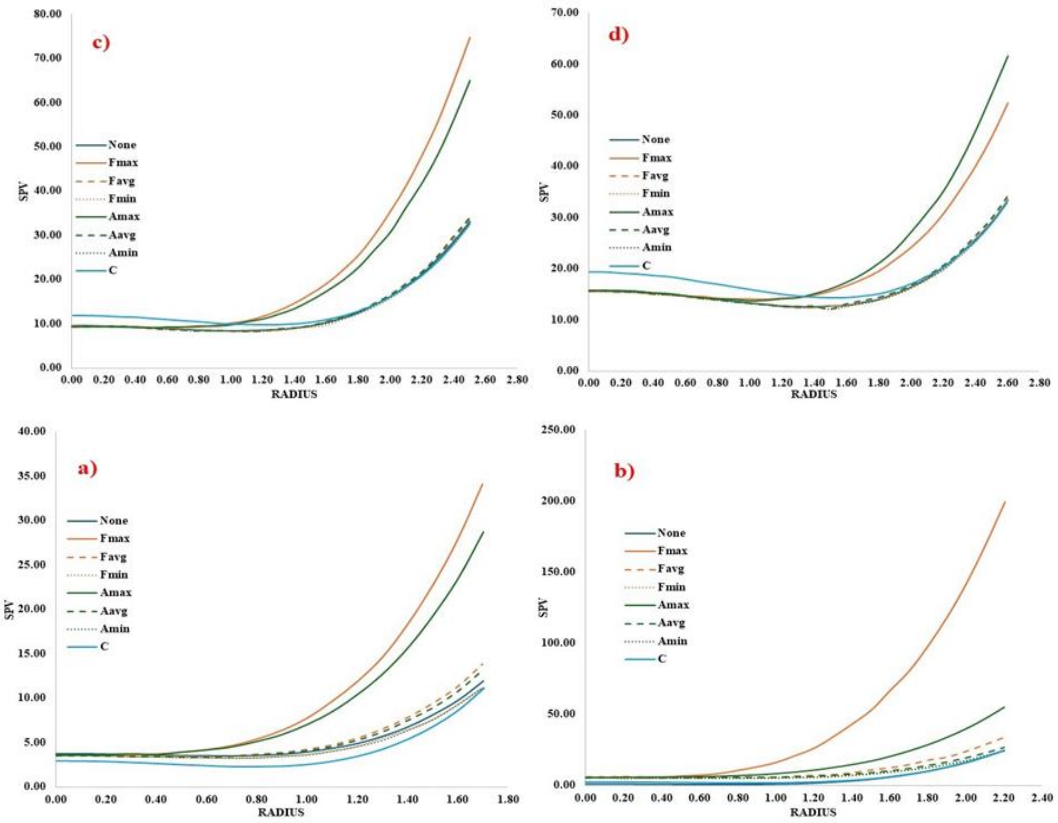

Figure 2: *(a) VDG for CCD K = 3. (b) VDG for CCD K = 5. (c) VDG for CCD K = 6. (d) VDG for CCD K = 7.*

IV. Interpretation of FDS of factors k = 2 to 7 of spherical $\sqrt{k}$ alpha value:

Figure 3 of *FDS* plot depicts that the impact of missing and non-missing observations is not significantly different across most factors. For factor k = 2, a missing center run must be distinct from all other factorial, axial, and non-missing design points. In contrast, the various observations for all other factors follow similar trajectories, staying close to each other within 75% to 85% of the design space region. When considering missing factorial and axial observations, it's notable that only for factor k = 7 does the axial observation have the highest *SPV*. The factorial observation has the highest *SPV* for all other factors, with factor k = 5 showing an exceptional *SPV* value of 191.86. A *G*-efficiency of 100% is achieved by all factors at various design points when the *FDS* region is approximately 80%. Beyond 85% of the *FDS* region, all factors experience a significant increase in the maximum *SPV*.

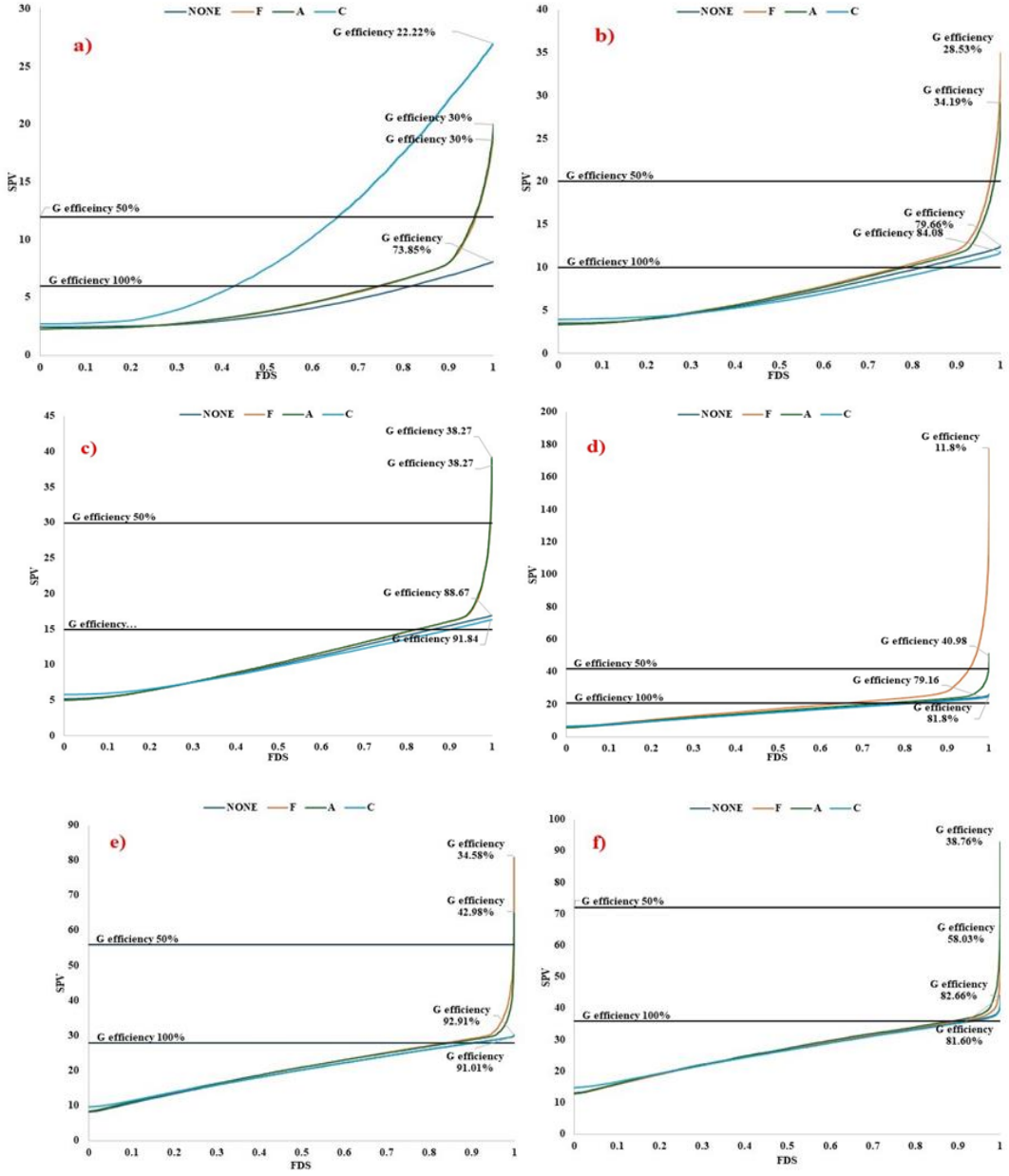

Figure 3: *(a) FDS for CCD K = 2. (b) FDS for CCD K = 3. (c) FDS for CCD K = 4. (d) FDS for CCD K = 5. (e) FDS for CCD K = 6. (F) FDS for CCD K = 7*

V. Interpretation of FDS of factors K = 3, 5, 6 and 7 of rotatable alpha value

As the factor increases (such as 5, 6, and 7), the coverage of the *FDS* region over 100% of the *G*-efficiency also increases. Specifically, for factor k = 5, both missing and non-missing observations reach the 100% *G*-efficiency line at 55% of the *FDS* region. For factor k = 6, they reach this line at 85% of the *FDS* region; for factor k = 7, they reach it at 95% of the *FDS* region

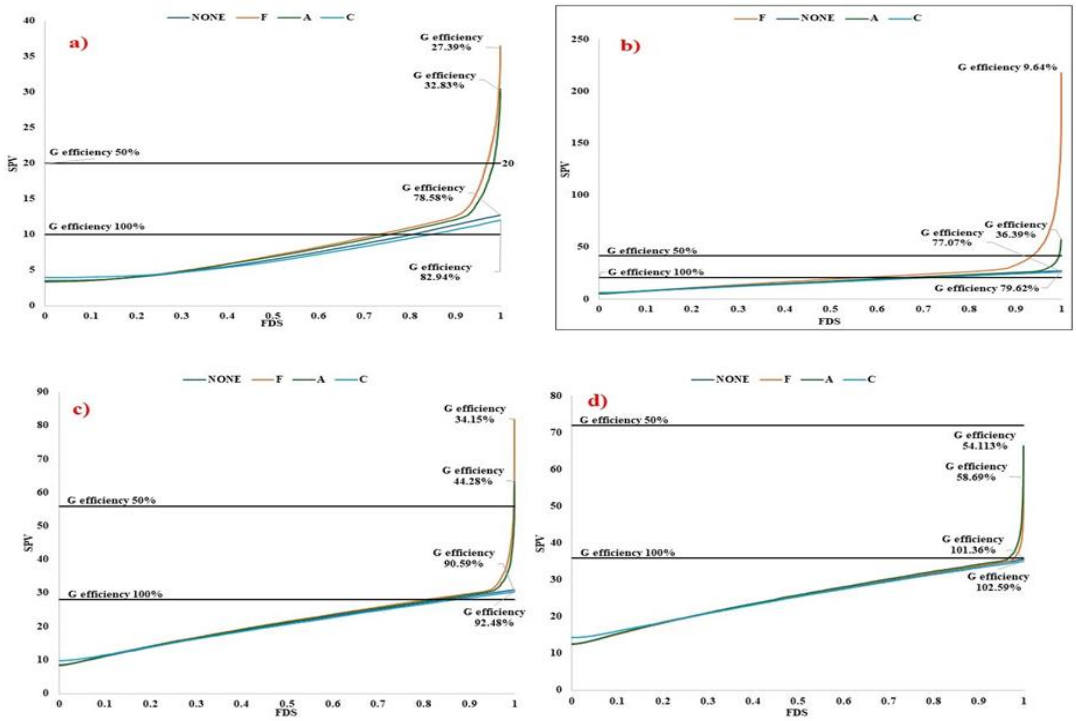

Figure 4: *(a) FDS for CCD K = 3. (b) FDS for CCD K = 5. (c) FDS for CCD K = 6. (d) FDS for CCD K = 7.*

IV. Conclusion

In an experimental setup, the absence of a single observation could potentially influence the results. There are diverse design points in a central composite design context, each with unique characteristics. In the context of a spherical region of interest or near a rotatable alpha for each factor, it's shown that missing a factorial observation influences the *SPV* value, leading to a maximum variance. Through *VDG*, we observe that the *SPV* value begins to rise approximately when the radius $r \geq 1$. On the other hand, in the *FDS* region, factorial data results occupying more than 85% of the fraction of the design space exhibit a significant SPV value. The average and minimum variances across all factors and various design points tend to remain close to each other. The results indicate that the absence of a center run, when compared to scenarios with no missing design points, does not influence the experiment's outcome within any spherical region of interest. The study highlights the benefits of employing graphical representations when dealing with missing observations in the region of interest. This approach can be expanded to include various second-order composite, computer-generated, and optimal designs to examine robust scenarios, making it a valuable tool in experimental design.

References

[1] Ahmad, T. Gilmour, S. G. and Muhammad Arshad, H. (2020). Comparisons of augmented pairs designs and subset designs. *Communications in Statistics: Simulation and Computation*, *49*(7), 1898–1921. https://doi.org/10.1080/03610918.2018.1508703

[1] Ahmad, T. Gilmour, S. G. and Muhammad Arshad, H. (2020). Comparisons of augmented pairs designs and subset designs. *Communications in Statistics: Simulation and Computation*, *49*(7), 1898–1921. https://doi.org/10.1080/03610918.2018.1508703

[2] Akhtar, M. and Prescott, P. (1986). Response surface designs robust to missing observations. *Communications in Statistics - Simulation and Computation*, *15*(2), 345–363. https://doi.org/10.1080/03610918608812512

[3] Alanazi, K. Georgiou, S. D. and Stylianou, S. (2023). On the robustness of central composite designs when missing two observations. *Quality and Reliability Engineering International*, *39*(4), 1143–1171. https://doi.org/10.1002/qre.3281

[4] Alrweili, H. Georgiou, S. and Stylianou, S. (2019). Robustness of response surface designs to missing data. *Quality and Reliability Engineering International*, *35*(5), 1288–1296. https://doi.org/10.1002/qre.2524

[5] Borkowski, J. J. (1995). Spherical prediction-variance properties of central composite and Box—Behnken designs. *Technometrics*, *37*(4), 399–410. https://doi.org/10.1080/00401706.1995.10484373

[6] Box, G. E. P. and Hunter, J. S. (1957). Multi-Factor Experimental Designs for Exploring Response Surfaces. *The Annals of Mathematical Statistics*, *28*(1), 195–241. https://doi.org/10.1214/aoms/1177707047

[7] Box, G. E. P. and Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(1), 1–38. https://doi.org/10.1111/j.2517-6161.1951.tb00067.x

[8] Draper, N. R. (1961). Missing Values in Response Surface Designs. *Technometrics*, *3*(3), 389. https://doi.org/10.2307/1266729

[9] Giovannitti-Jensen, A. and Myers, R. H. (1989). *American Society for Quality Graphical Assessment of the Prediction Capability of Response Surface Designs* (Vol. 31).

[10] Hayat, H. Akbar, A. Ahmad, T. Bhatti, S. H. and Ullah, M. I. (2023). Robustness to missing observation and optimalities of response surface designs with regular and complex structure. *Communications in Statistics - Simulation and Computation*, *52*(11), 5213–5230. https://doi.org/10.1080/03610918.2021.1977954

[11] Hemavathi, M. Varghese, E. Shekhar, S. Athulya, C. K. Ebeneezar, S. Gills, R. and Jaggi, S. (2022). Robustness of sequential third-order response surface design to missing observations. *Journal of Taibah University for Science*, *16*(1), 270–279. https://doi.org/10.1080/16583655.2022.2046398

[12] Khuri, A. I. and Mukhopadhyay, S. (2010, March). Response surface methodology. *Wiley Interdisciplinary Reviews: Computational Statistics*. https://doi.org/10.1002/wics.73

[13] Li, J. Li, L. Borror, 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 & Quantitative Management*, *6*(4), 433–449. https://doi.org/10.1080/16843703.2009.11673209

[14] On comparing the prediction variances of some central composite designs in spherical regions a review. (n.d.).

[15] Onwuamaeze, C. U. (2021). Optimal prediction variance properties of some central composite designs in the hypercube. *Communications in Statistics - Theory and Methods*, *50*(8), 1911–1924. https://doi.org/10.1080/03610926.2019.1656746

[16] Ozol-Godfrey, A. Anderson-Cook, C. M. and Montgomery, D. C. (2005). Fraction of Design Space Plots for Examining Model Robustness. *Journal of Quality Technology*, *37*(3), 223–235. https://doi.org/10.1080/00224065.2005.11980323

[17] Park, Y. J. Richardson, D. E. Montgomery, D. C. Ozol-Godfrey, A. Borror, C. M., & Anderson-Cook, C. M. (2005). Prediction variance properties of second-order designs for cuboidal regions. *Journal of Quality Technology*, *37*(4), 253–266. https://doi.org/10.1080/00224065.2005.11980329

[18] Vining, G. G. and Myers, R. H. (1991). A graphical approach for evaluating response surface designs in terms of the mean squared error of prediction. *Technometrics*, *33*(3), 315–326.

https://doi.org/10.1080/00401706.1991.10484837

[19] Zahran, A. Anderson-Cook, C. M. and Myers, R. H. (2003). Fraction of design space to assess prediction capability of response surface designs. *Journal of Quality Technology*, *35*(4), 377–386. https://doi.org/10.1080/00224065.2003.11980235