วารสารวิชาการพระจอมเกล้าพระนครเหนือ ปีที่ 32, ฉบับที่ 3 ก.ค.–ก.ย. 2565
The Journal of KMUTNB., Vol. 32, No. 3, Jul.–Sep. 2022

บทความวิจัย

การเปรียบเทียบวิธีประมาณค่าสูญหายในแผนแบบพื้นผิวตอบสนอง

ณัฐชยา รตะสุขารมย์* บุญอ้อม โฉมที จันทร์ธา วงษ์อู่ทอง และ สุดารัตน์ นิจสุนกิจ
สาขาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

* ผู้นิพนธ์ประสานงาน โทรศัพท์ 08 5355 6229 อีเมล: natchaya.rata@gmail.com DOI: 10.14416/j.kmutnb.2022.01.001
รับเมื่อ 17 กรกฎาคม 2563 แก้ไขเมื่อ 15 กันยายน 2563 ตอบรับเมื่อ 12 ตุลาคม 2563 เผยแพร่ออนไลน์ 7 มกราคม 2565

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสูญหายในแผนแบบพื้นผิวตอบสนอง 4 วิธี คือ วิธีค่าเฉลี่ย (Mean Imputation; MI) วิธีการถดถอย (Regression Imputation; RI) วิธีการถดถอยแบบสโตแคสติก (Stochastic Regression Imputation; SRI) และวิธีเคเนียร์เรสเนเบอร์ (K-Nearest Neighbor; KNN) ในแผนแบบพื้นผิวตอบสนอง 4 แผนแบบ ได้แก่ แผนแบบเซ็นทรัลคอมโพสิต (Central Composite Design; CCD) แผนแบบสมอลคอมโพสิต (Small Composite Design; SCD) แผนแบบบ็อกซ์-เบห์นเคน (Box-Behnken Design; BBD) และแผนแบบไฮบริด (Hybrid Design) ภายในขอบเขตทรงกลม เมื่อมีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย กำหนดความแปรปรวนของความคลาดเคลื่อนเท่ากับ 0.5, 1 และ 1.5 เกณฑ์การเปรียบเทียบพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error; MSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error; MAE) ผลการศึกษาพบว่า ในแผนแบบขนาดใหญ่ (N = 26, 27) วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดี แผนแบบขนาดกลาง ($16 \le N \le 19$) โดยส่วนใหญ่วิธี RI และ MI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดี ส่วนแผนแบบขนาดเล็ก ($12 \le N \le 14$) โดยส่วนใหญ่วิธี MI และวิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีทุกวิธีจะมีประสิทธิภาพเพิ่มขึ้น เมื่อจำนวนการทำซ้ำที่จุดศูนย์กลางของแผนแบบ (n_c) เพิ่มขึ้น และเมื่อความแปรปรวนของความคลาดเคลื่อนลดลง

คำสำคัญ: แผนแบบพื้นผิวตอบสนอง การประมาณค่าสูญหาย ความคลาดเคลื่อนกำลังสองเฉลี่ย ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

การอ้างอิงบทความ: ณัฐชยา รตะสุขารมย์, บุญอ้อม โฉมที, จันทร์ธา วงษ์อู่ทอง, และ สุดารัตน์ นิจสุนกิจ, "การเปรียบเทียบวิธีประมาณค่าสูญหายในแผนแบบพื้นผิวตอบสนอง," *วารสารวิชาการพระจอมเกล้าพระนครเหนือ*, ปีที่ 32, ฉบับที่ 3, หน้า 758–769, ก.ค.–ก.ย. 2565.

Research Article

A Comparison of Missing Value Estimation Methods for Response Surface Design

Natchaya Ratasukharom*, Boonorm Chomtee, Chantha Wongoutong and Sudarat Nidsunkid
Department of Statistics, Faculty of Science, Kasetsart University, Bangkok, Thailand

* Corresponding Author, Tel. 08 5355 6229, E–mail: natchaya.rata@gmail.com
DOI: 10.14416/j.kmutnb.2022.01.001
Received 17 July 2020; Revised 15 September 2020; Accepted 12 October 2020; Published online: 7 January 2022

Abstract

The objective of this research is to compare efficiency of missing value estimation methods in response surface designs. The missing value estimation methods considered in the research are the four imputation methods: Mean Imputation (MI), Regression Imputation (RI), Stochastic Regression Imputation (SRI) and K-nearest Neighbor imputation (KNN). The four response surface designs in a spherical region with 3 and 4 design variables (k = 3, 4): Central Composite Design (CCD), Small Composite Design (SCD), Box-behnken Design (BBD) and Hybrid design are used in this study. The variance of errors are 0.5, 1 and 1.5. The criteria for comparing the efficiency are Mean Square Error (MSE) and Mean Absolute Error (MAE). The results show that the RI method performs the best in the large design sizes (N = 26, 27). The RI and MI methods mostly performs well in the medium design sizes ($16 \leq N \leq 19$). The MI and KNN methods mostly perform well in the small design sizes ($12 \leq N \leq 14$). The efficiency of all imputation methods increases when center points (n_c) increases and the variance of error decreases.

Keywords: Response Surface Design, Missing Value Estimation, Mean Square Error, Mean Absolute Error

Please cite this article as: N. Ratasukharom, B. Chomtee, C. Wongoutong, and S. Nidsunkid, "A comparison of missing value estimation methods for response surface design," *The Journal of KMUTNB*, vol. 32, no. 3, pp. 758–769, Jul.–Sep. 2022 (in Thai).

1. บทนำ

การออกแบบการทดลองเป็นการค้นหาคำตอบของการทดลองด้วยวิธีที่มีแบบแผนเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระ (Independent Variable) หรือปัจจัยที่ใช้ในการทดลองกับค่าผลตอบสนอง (Response Variable) ที่ได้จากการทดลอง ซึ่งปัจจุบันได้มีการนำการออกแบบการทดลองโดยวิธีพื้นผิวตอบสนอง (Response Surface Methodology; RSM) มาใช้อย่างแพร่หลาย เช่น ด้านการเกษตร ด้านการแพทย์ ด้านอุตสาหกรรม เนื่องจากเป็นวิธีที่รวมเทคนิคทางด้านคณิตศาสตร์ และสถิติมาใช้ในการสร้างตัวแบบ และวิเคราะห์ปัญหา ในกรณีที่ผลตอบสนองมีความสัมพันธ์กับปัจจัยหลายตัว เพื่อค้นหาระดับของปัจจัยที่ทำให้ผลตอบสนองมีค่าที่ดีที่สุด นั่นก็คือการหาระดับของปัจจัยที่เหมาะสม ที่ก่อให้เกิดผลตอบสนองที่มากที่สุดหรือน้อยที่สุดตามที่ผู้วิจัยต้องการ แต่ในบางครั้งของการทดลอง แม้จะมีการวางแผนการทดลองเป็นอย่างดีแล้วแต่ก็มีโอกาสที่จะเกิดการสูญหายของข้อมูล หรืออาจพบว่า ค่าผลตอบสนองที่เป็นผลลัพธ์ของการทดลองนั้นไม่ถูกต้องซึ่งจะเรียกค่าผลตอบสนองดังกล่าวว่า “ข้อมูลสูญหาย (Missing Value)” การหายไปของค่าผลตอบสนองอาจทำให้ผลลัพธ์ของการทดลองเกิดความผิดพลาดได้ โดยเฉพาะอย่างยิ่งในกรณีของการทดลองแบบไม่มีการทำซ้ำหรือในการทดลองที่มีค่าใช้จ่ายค่อนข้างสูง และในการทดลองที่ต้องใช้ระยะเวลาในการทดลองที่ยาวนาน ซึ่งยากต่อการทำการทดลองใหม่ จึงต้องมีการประมาณค่าข้อมูลสูญหาย เพื่อจะได้ผลลัพธ์ของการทดลองที่แม่นยำและมีประสิทธิภาพ โดยแผนแบบการทดลองที่นิยมนำมาใช้เป็นอย่างมาก ได้แก่ แผนแบบ Central Composite Design (CCD), Small Composite Design (SCD), Box-Behnken Design (BBD) และ Hybrid Design ดังนั้นในการศึกษานี้ ผู้วิจัยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าข้อมูลสูญหาย 4 วิธี ในแผนแบบพื้นผิวตอบสนองเมื่อมีค่าข้อมูลสูญหาย 1 ค่า ภายใต้ขอบเขตทรงกลม คือ วิธีค่าเฉลี่ย (Mean Imputation; MI), วิธีการถดถอย (Regression Imputation; RI), วิธีการถดถอยแบบสโตแคสติก (Stochastic Regression Imputation; SRI) และวิธีเค-เนียร์เรสเนเบอร์ (K-Nearest Neighbor; KNN) ซึ่งจะเป็นประโยชน์กับผู้ที่ทำการทดลองในแผนแบบพื้นผิวตอบสนอง เมื่อมีข้อมูลสูญหายเกิดขึ้นระหว่างทำการทดลอง และผู้ทดลองจะสามารถเลือกใช้วิธีการประมาณค่าข้อมูลสูญหายที่เหมาะสมกับแต่ละแผนแบบการทดลองได้

2. วัสดุ อุปกรณ์และวิธีการวิจัย

งานวิจัยครั้งนี้ทำการศึกษาด้วยข้อมูลจำลองโดยวิธีมอนติคาร์โล ในแต่ละสถานการณ์ทำซ้ำ 15,000 ครั้ง ด้วยโปรแกรม MATLAB สำหรับแผนแบบพื้นผิวตอบสนอง CCD, SCD, BBD และ Hybrid Design ที่มีจำนวนปัจจัย (k) เท่ากับ 3 และ 4 ปัจจัย มีจำนวนการทำซ้ำที่ Center Points (n_c) เท่ากับ 2 ถึง 3 การทำซ้ำ

2.1 แผนแบบที่ศึกษา

2.1.1 Central Composite Design (CCD)

แผนแบบ CCD [1] ประกอบด้วย 3 ส่วน ดังนี้

1) Factorial Points เป็นการนำแผนการทดลองแบบแฟกทอเรียลจำนวน k ปัจจัย ซึ่งแต่ละปัจจัยมี 2 ระดับ คือ ที่ระดับ +1 และ −1 จะมีจำนวน 2^k จุด

2) Axial Points เป็นการกำหนดค่าให้ปัจจัยใดปัจจัยหนึ่งเป็น $\sqrt{k}$ ในขณะที่ปัจจัยอื่นมีค่าอยู่ที่ระดับกลาง (หรือค่า 0) จะมีจำนวน $2k$ จุด

3) Center Points เป็นการกำหนดค่าปัจจัยทั้งหมดให้อยู่ที่ระดับกลาง (หรือค่า 0) จะมีจำนวน n_c จุด

ดังนั้น ขนาดของแผนแบบ (N) เกิดจากการรวมกันของทั้ง 3 ส่วน คือ Factorial Points, Axial Points และ Center Points นั่นคือ $N = 2^k + 2k + n_c$

ตัวอย่าง แผนแบบ CCD, $k = 3, N_c = 2$ ภายใต้ขอบเขตทรงกลม มีลักษณะดังนี้

$$\boldsymbol{M}_1 = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \\ -1.732 & 0 & 0 \\ 1.732 & 0 & 0 \\ 0 & -1.732 & 0 \\ 0 & 1.732 & 0 \\ 0 & 0 & -1.732 \\ 0 & 0 & 1.732 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

จากเมทริกซ์ แถวที่ 1 ถึง 8 เป็นส่วนของ Factorial Points แถวที่ 9 ถึง 14 เป็นส่วนของ Axial Points และแถวที่ 15 ถึง 16 เป็นส่วนของ Center Points

2.1.2 Small Composite Designs (SCD)

แผนแบบ SCD [2] มีแนวคิดมาจากแผนแบบ CCD ที่มีหน่วยทดลองค่อนข้างจำกัดจึงทำให้ในส่วนของ Factorial Points จะไม่เป็น Full Factorial แต่จะมีจำนวนเท่ากับ 2^{k-p} แฟรกชันนอลแฟกทอเรียล ซึ่งในการศึกษานี้กำหนดให้ $p = 1$

ตัวอย่าง แผนแบบ SCD, $k = 3, N_c = 2$ ภายใต้ขอบเขตทรงกลม มีลักษณะดังนี้

$$\boldsymbol{M}_2 = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ -1.732 & 0 & 0 \\ 1\ 732 & 0 & 0 \\ 0 & -1.732 & 0 \\ 0 & 1.732 & 0 \\ 0 & 0 & -1\ 732 \\ 0 & 0 & 1.732 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

จากเมทริกซ์ จะเห็นได้ว่าในส่วนของ Factorial Points มีจำนวนลดลงเมื่อเทียบกับแผนแบบ CCD โดยในส่วนของ Factorial Points สร้างจากแผนแบบ $2^{3-1} = 2^2 = 4$ แฟรกชันนอลแฟกทอเรียลที่มี Defining Relation $I = -x_1 x_2 x_3$

2.1.3 Box-Behnken Design (BBD)

แผนแบบ BBD [3] เป็นแผนแบบสำหรับ $k \geq 3$ โดยในแต่ละปัจจัยมี 3 ระดับ (Three-level Design) สร้างขึ้นจากแผนแบบบล็อกไม่สมบูรณ์สมดุล (Balanced Incomplete Block Design; BIBD) และแผนแบบ 2^k แฟกทอเรียล โดยกำหนดพารามิเตอร์ที่ใช้ ดังนี้ k คือ จำนวนปัจจัย b คือ จำนวนบล็อกในแผนแบบบล็อกไม่สมบูรณ์ t คือ จำนวนปัจจัยที่ปรากฏต่อบล็อก r คือ จำนวนบล็อกที่ปรากฏในแต่ละปัจจัย λ คือ จำนวนครั้งที่ปัจจัยแต่ละคู่ปรากฏในบล็อกเดียวกัน โดย $\lambda = \frac{r(t-1)}{k-1}$ BBD สร้างจาก 1) ปัจจัย t ที่ปรากฏในแต่ละบล็อกใน BIBD จะถูกแทนที่ด้วย t คอลัมน์ซึ่งเป็นส่วนประกอบของ 2^k แฟกทอเรียล ที่มีระดับของปัจจัย ±1 2) คอลัมน์ที่เหลือ $k - 1$ คอลัมน์ ถูกกำหนดในมีระดับของปัจจัยเท่ากับ 0 3) จุดศูนย์กลางของแผนแบบอยู่ที่ $(x_1, x_2, ..., x_3) = (0,0,...,0)$

ตัวอย่าง แผนแบบ BBD, $k = 3, N_c = 2$ ภายใต้ขอบเขตทรงกลม มีลักษณะดังนี้

$$\boldsymbol{M}_3 = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} -1.2247 & -1\ 2247 & 0 \\ -1.2247 & 1.2247 & 0 \\ 1.2247 & -1\ 2247 & 0 \\ 1.2247 & 1.2247 & 0 \\ -1\ 2247 & 0 & -1\ 2247 \\ -1\ 2247 & 0 & 1\ 2247 \\ 1.2247 & 0 & -1\ 2247 \\ 1\ 2247 & 0 & 1.2247 \\ 0 & -1.2247 & -1.2247 \\ 0 & -1\ 2247 & 1\ 2247 \\ 0 & 1.2247 & -1.2247 \\ 0 & 1.2247 & 1.2247 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

แผนแบบ $\boldsymbol{M_3}$ สร้างจาก BIBD ที่มี $k = 3$, $\lambda = \frac{3(2-1)}{3-1} = \frac{3}{2}$

2.1.4 Hybrid Design

เป็นแผนแบบสำหรับ $k \geq 3$ มีแนวคิดมาจากแผนแบบ CCD ที่มี $k-1$ ปัจจัย ส่วนปัจจัยที่เหลือจะถูกกำหนดให้มี 4 ระดับของปัจจัย [4]

ตัวอย่าง แผนแบบ Hybrid H310, $k=3, N_c=2$ ภายใต้ขอบเขตทรงกลม มีลักษณะดังนี้

$$\boldsymbol{M}_{H310} = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 1.2960 \\ 0 & 0 & -0.1360 \\ -1 & -1 & 0\ 6386 \\ 1 & -1 & 0\ 6386 \\ -1 & 1 & 0.6386 \\ 1 & 1 & 0.6386 \\ 1.7360 & 0 & -0.9273 \\ -1.7360 & 0 & -0.9273 \\ 0 & -1.7360 & -0.9273 \\ 0 & 1.7360 & -0.9273 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

จากเมทริกซ์ $\boldsymbol{M}_{H310}$ ประกอบด้วยแผนแบบ CCD ที่มี $k-1=3-1=2$ ตัวแปร ในปัจจัย x_1 และ x_2 ส่วนปัจจัย x_3 ถูกกำหนดให้มี 4 ระดับของปัจจัย

1) กำหนดค่าของตัวแปรปัจจัย (x) ตามลักษณะของแผนแบบ CCD, SCD, BBD และ Hybrid (H310, H311A, H311B, H416A, H416B และ H416C) ในขอบเขตทรงกลม

2) กำหนดค่าความคลาดเคลื่อนสุ่ม (ε) ให้มีการแจกแจงปรกติที่มีค่าเฉลี่ยเท่ากับ 0 ความแปรปรวนเท่ากับ 0.5, 1 และ 1.5

3) สร้างค่าตัวแปรผลตอบสนองที่มีความสัมพันธ์กับตัวแปรปัจจัย (x) โดยใช้ตัวแบบกำลังสองเต็ม (Full Second Order Model) ดังสมการที่ (1)

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

เมื่อ y คือ ตัวแปรผลตอบสนอง x_i คือ ปัจจัยที่ i ; $i = 1,2,...,k$ k คือ จำนวนปัจจัย β_0 คือ จุดตัดแกน y (y Intercept) $\Sigma_{i=1}^{k}\beta_i x_i$ คือ เทอมของอิทธิพลเชิงเส้น (Linear Effect) $\Sigma_{i=1}^{k}\beta_{ii}x_i^2$ คือ เทอมของอิทธิพลกำลังสอง (Quadratic Effect) $\Sigma_{i=1}^{k-1}\Sigma_{j=i+1}^{k}\beta_{ij}x_i x_j$ คือ เทอมของอิทธิพลร่วม (Cross-product Effect) และ ε คือ ความคลาดเคลื่อนสุ่ม ซึ่ง $\varepsilon \sim N(0,\sigma^2)$ โดยที่ [5] ได้แนะนำไว้ว่า เพื่อควบคุมความแปรปรวนของตัวแปรผลตอบสนองให้มีขนาดเท่ากัน ควรกำหนดให้สัมประสิทธิ์การถดถอยมีค่าอยู่ระหว่าง –1 ถึง 1 ดังนั้นในการศึกษานี้จึงกำหนดค่า $\beta_0 = \beta_{ii} = \beta_{ij} = 1$

4) สร้างค่าตัวแปรตอบสนองภายใต้ขอบเขตทรงกลม เมื่อ $k=3$ สำหรับแผนแบบ CCD, SCD, BBD, H310, H311A, H311B และ $k=4$ สำหรับแผนแบบ CCD, SCD, BBD, H416A, H416B และ H416C โดยกำหนดให้แผนแบบมีรัศมี เท่ากับ $\alpha = \sqrt{k}$ และกำหนดจำนวนจุดศูนย์กลางของแผนแบบ (n_c) เท่ากับ 2 และ 3 ในงานวิจัยนี้จะกำหนดขนาดของแผนแบบ (N) โดยแผนแบบขนาดเล็กมี $12 \leq N \leq 14$ ขนาดกลางจะมี $16 \leq N \leq 19$ และขนาดใหญ่จะมี $N = 26, 27$

2.1.5 การประมาณค่าสูญหาย มี 4 วิธี ดังนี้

1) วิธีค่าเฉลี่ย (MI) เป็นวิธีแทนค่าสูญหายด้วยค่าคงที่ [6] โดยประมาณค่าข้อมูลสูญหายด้วยค่าเฉลี่ยของตัวแปรเดียวกันจากข้อมูลที่เก็บมาได้ โดยสำหรับแผนแบบ CCD และ SCD จะทำการประมาณค่าสูญหายโดยการแทนค่าข้อมูลสูญหายด้วยการใช้ค่าเฉลี่ยแบบแบ่งชั้น (Class Mean Imputation) โดยจะทำการพิจารณาก่อนว่าข้อมูลที่สูญหายนั้นเกิดขึ้นที่ส่วนใดของแผนแบบ (ส่วน Factorial Points ส่วน Axial Points หรือส่วน Center Points) จากนั้นจะประมาณค่าสูญหาย โดยนำค่าของข้อมูลที่มีอยู่ หรือข้อมูลที่ไม่สูญหายในแต่ละส่วนมาคำนวณค่าเฉลี่ย ส่วนแผนแบบ BBD และ Hybrid จะทำการประมาณค่าสูญหายโดยใช้ค่าเฉลี่ยของข้อมูลที่ไม่สูญหาย ดังสมการที่ (2)

$$\hat{y}_m = \frac{\sum_{i=1}^{N-1} y_i}{N-1} \quad (2)$$

2) วิธีการถดถอย (RI) เป็นวิธีที่นำข้อมูลที่มีอยู่มาศึกษา

ความสัมพันธ์เชิงเส้นเพื่อสร้างสมการถดถอย โดยจะใช้สมการกำลังสอง (Second-order Model) ประมาณค่าสูญหายด้วยชุดข้อมูลที่มีอยู่ (x_i,y_i) ; $i = 1, 2,\ldots, N-1$ มาคำนวณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด (Ordinary Least Square Method; OLS) เพื่อประมาณค่าสูญหายของตัวแปรตอบสนอง ($\hat{y}_m$) [7] โดยมีขั้นตอนดังนี้

2.1) คำนวณค่าสัมประสิทธิ์การถดถอยจากข้อมูลที่ไม่สูญหาย (x_i,y_i) ด้วยวิธีกำลังสองน้อยที่สุด ได้ค่าประมาณสัมประสิทธิ์การถดถอย ดังสมการที่ (3)

$$\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^{*'}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*'}\boldsymbol{y}^* \tag{3}$$

เมื่อ $\boldsymbol{X}^*$ และ $\boldsymbol{y}^*$ เป็นชุดข้อมูลที่มีอยู่ของ $\boldsymbol{X}$ และ $\boldsymbol{Y}$

2.2) นำค่าประมาณสัมประสิทธิ์การถดถอย ($\hat{\boldsymbol{\beta}}^*$) ที่ได้จากขั้นตอนที่ 1 มาประมาณค่าสูญหายของตัวแปรผลตอบสนอง โดยพิจารณาจากสมการถดถอยเชิงเส้นกำลังสอง ดังสมการที่ (4)

$$\hat{y}_m = \boldsymbol{X}_m\hat{\boldsymbol{\beta}}^* \tag{4}$$

เมื่อ $\hat{y}_m$ เป็น ค่าประมาณของค่าสูญหาย
$X_m = [1\ x_{m1}\cdots x_{mk}\ \ x_{m1}x_{m2}\cdots x_{m(k-1)}x_{mk}\ \ x_{m1}^2\cdots x_{mk}^2]$ โดยที่ $x_{m1}\cdots x_{mk}$ เป็น Linear Term $x_{m1}x_{m2}\cdots x_{m(k-1)}x_{mk}$ เป็น Cross-product Term และ $x_{m1}^2\cdots x_{mk}^2$ เป็น Quadratic Term ในตัวแบบที่มีข้อมูลสูญหาย

3) วิธีการถดถอยแบบสโตแคสติก (SRI) [8] เป็นการประมาณค่า y_m จากข้อมูล (x_i,y_i) ; $i = 1, 2,\ldots, N-1$ ที่มีอยู่ โดยประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด (OLS) เช่นเดียวกับวิธี RI แต่จะเพิ่มเทอมของความคลาดเคลื่อนสุ่มเข้ามาในสมการถดถอย ดังสมการที่ (5)

$$\hat{y}_m = \boldsymbol{X}_m\hat{\boldsymbol{\beta}}^* + \hat{\varepsilon}_m \tag{5}$$

เมื่อ $\hat{y}_m$ เป็น ค่าประมาณของค่าสูญหาย และ $\hat{\varepsilon}_m$ เป็นค่าประมาณของความคลาดเคลื่อนสุ่มของค่าสูญหาย

4) วิธีเคเนียร์เรสเนเบอร์ (KNN) เป็นวิธีประมาณค่าสูญหายด้วยค่าเฉลี่ยจากข้อมูลที่มีความใกล้เคียงกับข้อมูลสูญหายมากที่สุดจำนวน K ตัว โดยจะพิจารณาจากระยะห่างยุคลิด (Euclidean Distance; D_{iN}) ระหว่างปัจจัยที่ไม่มีข้อมูลสูญหายกับปัจจัยที่มีข้อมูลสูญหาย [9] ซึ่งในที่นี้ให้เป็นปัจจัยที่ N สามารถคำนวณได้ ดังสมการที่ (6)

$$D_{iN} = \sqrt{\sum_{p=1}^{k}(x_{ip} - x_{Np})^2} \tag{6}$$

เมื่อ $i = 1, 2,\ldots, N-1$ โดยมีขั้นตอนดังนี้

4.1) พิจารณาค่า K ที่เป็นจำนวนเต็มที่มีค่าใกล้เคียงกับรากที่สองของจำนวนข้อมูลที่มีอยู่ นั่นคือ $K \approx \sqrt{N-1}$ เมื่อ $N-1$ เป็นจำนวนข้อมูลที่สมบูรณ์

4.2) พิจารณาระยะห่างยุคลิด (D_{iN}) ระหว่างปัจจัยที่มีค่าต่ำที่สุดจำนวน K ตัว

4.3) คำนวนค่าเฉลี่ยของตัวแปรตอบสนองที่สอดคล้องกับ D_{iN} ที่ต่ำที่สุด จำนวน K ตัว

2.1.6 เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสูญหายจะพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Square Error; MSE) และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error; MAE) จากการทำซ้ำ 15,000 รอบ (t) ของแต่ละสถานการณ์สำหรับวิธีการประมาณค่าที่ศึกษา โดยวิธีใดให้ค่า MSE และ MAE ต่ำที่สุด วิธีนั้นจะมีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด คำนวณดังสมการที่ (7) และ (8) ดังนี้

$$MSE = \frac{1}{t}\sum_{m=1}^{N}(y_m - \hat{y}_m)^2 \tag{7}$$

$$MAE = \frac{1}{t}\sum_{m=1}^{N}|y_m - \hat{y}_m| \tag{8}$$

เมื่อ y_m คือ ค่าจริงของค่าตอบสนอง, $\hat{y}_m$ คือ ค่าประมาณของค่าตอบสนอง และ t คือ จำนวนรอบที่ทำซ้ำ

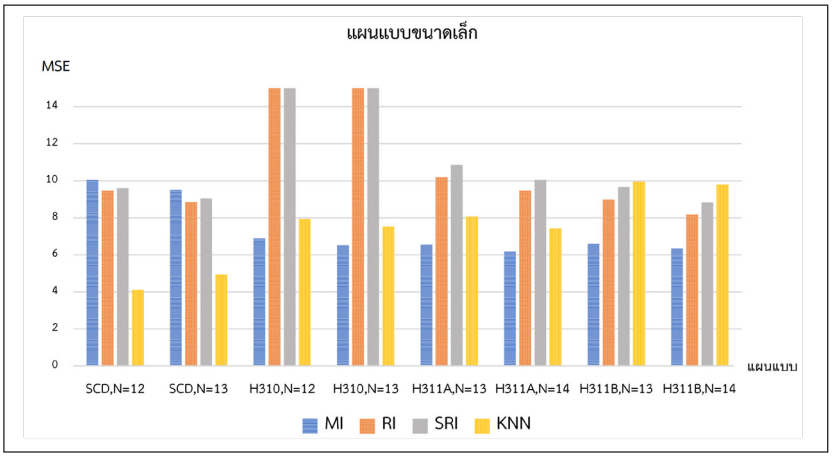
3. ผลการทดลอง

เมื่อความแปรปรวนของความคลาดเคลื่อนเท่ากับ 0.5 พบว่า ในแผนแบบ CCD และ BBD ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด ส่วนแผนแบบ SCD วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีเมื่อแผนแบบมีจำนวนปัจจัยเท่ากับ 3 ปัจจัย ส่วนใน 4 ปัจจัย วิธี RI จะมีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด และในแผนแบบ H310, H416B และ H416C วิธี MI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด ส่วนในแผนแบบ H311A, H311B และ H416A วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด

เมื่อความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1 พบว่า ในแผนแบบ CCD และ BBD ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด สำหรับแผนแบบ SCD ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด ส่วนในแผนแบบ Hybrid ไม่ว่าจะเป็น H310, H311A, H311B, H416A, H416B และ H416C วิธี MI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด

เมื่อความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1.5 พบว่า ในแผนแบบ CCD ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด ส่วนแผนแบบ SCD ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด และในแผนแบบ BBD ที่มีจำนวนปัจจัยเท่ากับ 3 ปัจจัย วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด ส่วนใน 4 ปัจจัย วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด อย่างไรก็ตาม ในแผนแบบ Hybrid ใน H310, H311A, H311B, H416A, H416B และ H416C วิธี MI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด เช่นเดียวกันกับกรณีที่ความแปรปรวนของความคลาดเคลื่อนมีค่าเป็น 1

หากพิจารณาประสิทธิภาพของวิธีการประมาณค่าสูญหายจากขนาดของแผนแบบ (N) จะได้ผลการศึกษาดังนี้ แผนแบบขนาดเล็ก ($12 \leq N \leq 14$) พบว่า แผนแบบ SCD ที่มีจำนวนปัจจัยเท่ากับ 3 ปัจจัย วิธี KNN จะมีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด ส่วนในแผนแบบ H310, H311A และ 311B ที่มีจำนวนปัจจัยเท่ากับ 3 ปัจจัย ส่วนใหญ่วิธี MI จะมีประสิทธิภาพในการประมาณค่าสูญหายได้ดี ยกเว้นในแผนแบบ H311A และ H311B ที่มีค่าความแปรปรวนของความคลาดเคลื่อนเท่ากับ 0.5 วิธี RI จะมีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด แสดงดังตารางที่ 1 และรูปที่ 1



รูปที่ 1 ค่า MSE ของวิธีประมาณค่าสูญหายทั้ง 4 วิธี ในแผนแบบพื้นผิวตอบสนองขนาดเล็ก ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1

ตารางที่ 1 ค่า MSE และ MAE ในแผนแบบพื้นผิวตอบสนองขนาดเล็ก ($12 \leq N \leq 14$) เมื่อ $k = 3$

แผนแบบ	k	σ^2	N	เกณฑ์ที่ใช้เปรียบเทียบ							
				MSE				MAE			
				MI	RI	SRI	KNN	MI	RI	SRI	KNN
SCD	3	0.5	12	9.77	4.98	5.05	**3.79**	2.47	1.69	1.71	**1.58**
			13	8.93	4.63	4.76	**4.61**	2.31	1.62	1.66	**1.59**
		1	12	10.06	9.48	9.60	**4.11**	2.50	2.33	2.35	**1.63**
			13	9.51	8.86	9.05	**4.95**	2.37	2.20	2.24	**1.71**
		1.5	12	10.54	13.22	13.41	**4.51**	2.54	2.75	2.78	1.69
			13	9.64	12.38	12.67	**5.13**	2.36	2.58	2.63	1.73
H310	3	0.5	12	**6.55**	1434.72	1435.44	7.74	**1.73**	25.95	25.99	1.99
			13	**6.29**	1160.94	1161.36	7.51	**1.70**	22.62	22.67	1.92
		1	12	**6.89**	2448.24	2448.34	7.94	**1.82**	33.76	33.81	2.00
			13	**6.52**	1992.44	1993.40	7.52	**1.78**	29.25	29.32	1.90
		1.5	12	**7.10**	3333.89	3334.65	7.97	**1.90**	39.15	39.22	2.01
			13	**6.66**	2724.73	2724.80	7.55	**1.82**	34.44	34.51	1.92
H311A	3	0.5	13	8.67	**1.83**	2.28	3.36	2.37	**1.07**	1.20	1.55
			14	6.18	**4.83**	5.21	7.66	1.75	**1.58**	1.68	2.00
		1	13	**6.54**	10.19	10.86	8.07	**1.82**	2.33	2.45	2.06
			14	**6.19**	9.48	10.05	7.43	**1.78**	2.21	2.33	1.95
		1.5	13	**6.63**	14.63	15.48	7.94	**1.86**	2.79	2.93	2.03
			14	**6.49**	13.60	14.44	7.61	**1.84**	2.63	2.77	1.97
H311B	3	0.5	13	6.31	**4.63**	5.00	9.83	2.01	**1.60**	1.69	2.48
			14	6.34	**4.12**	4.48	10.29	2.00	**1.49**	1.59	2.42
		1	13	**6.60**	8.99	9.66	9.95	**2.05**	2.22	2.34	2.49
			14	**6.34**	8.17	8.82	9.79	**2.00**	2.08	2.21	2.35
		1.5	13	**6.75**	12.79	13.67	9.78	**2.07**	2.64	2.77	2.46
			14	**6.38**	11.68	12.57	9.43	**2.00**	2.48	2.62	2.30

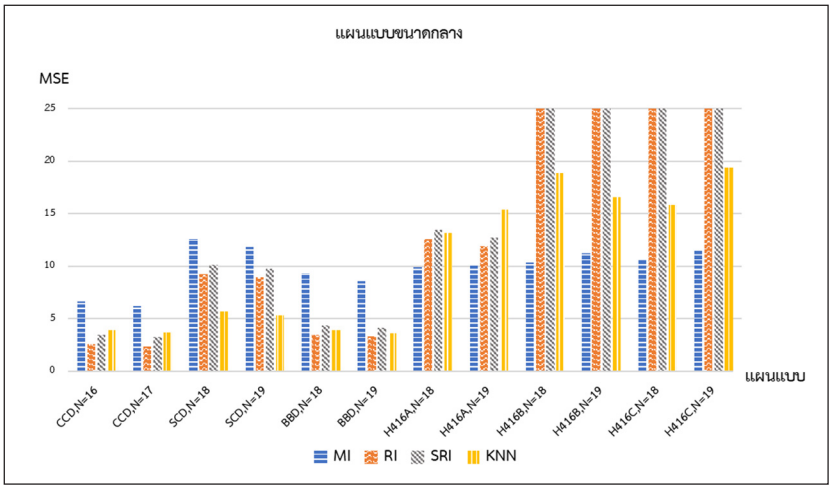
หมายเหตุ: ตัวหนา หมายถึงค่า MSE และ MAE ที่ต่ำที่สุด

ในแผนแบบขนาดกลาง ($16 \leq N \leq 19$) พบว่า ในแผนแบบ CCD ที่ $k = 3$ วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด แผนแบบ SCD ที่ $k = 4$ วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด ยกเว้นเมื่อความแปรปรวนเท่ากับ 0.5 พบว่า วิธี RI จะมีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด แผนแบบ BBD ที่ $k = 3$ วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด ยกเว้นเมื่อความแปรปรวนเท่ากับ 1.5 พบว่า วิธี KNN จะสามารถประมาณค่าสูญหายได้ดีที่สุด ส่วนในแผนแบบ H416A, H416B และ H416C ที่ $k = 4$ วิธี MI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุดยกเว้นในแผนแบบ H416A ที่มีความแปรปรวนเท่ากับ 0.5 พบว่า วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด สามารถแสดงได้ดังตารางที่ 2 และรูปที่ 2

สำหรับพื้นผิวตอบสนองขนาดใหญ่ ($N = 26, 27$) ในแผนแบบ CCD และ BBD ที่มีจำนวนปัจจัยเท่ากับ 4 ปัจจัย พบว่า วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด แสดงดังตารางที่ 3 และรูปที่ 3

ตารางที่ 2 ค่า MSE และ MAE ในแผนแบบพื้นผิวตอบสนองขนาดกลาง ($16 \leq N \leq 19$) เมื่อ $k = 3, 4$

แผนแบบ	k	σ^2	N	เกณฑ์ที่ใช้เปรียบเทียบ							
				MSE				MAE			
				MI	RI	SRI	KNN	MI	RI	SRI	KNN
CCD	3	0.5	16	6.27	**1.32**	1.79	3.50	1.90	**0.91**	1.06	1.33
			17	5.85	**1.25**	1.70	3.30	1.79	**0.88**	1.03	1.29
		1	16	6.65	**2.56**	3.45	3.88	1.96	**1.27**	1.47	1.45
			17	6.24	**2.39**	3.28	3.72	1.87	**1.21**	1.43	1.42
		1.5	16	7.09	**3.67**	4.92	4.41	2.03	**1.52**	1.76	1.57
			17	6.65	**3.44**	4.75	4.16	1.94	**1.46**	1.72	1.53
SCD	4	0.5	18	12.10	**4.78**	5.23	5.14	2.61	**1.43**	1.75	1.66
			19	11.19	**4.54**	4.95	4.63	2.48	**1.34**	1.69	1.60
		1	18	12.58	9.28	10.12	**5.69**	2.63	2.32	2.45	**1.57**
			19	11.86	8.98	9.74	**5.30**	2.51	2.24	2.37	**1.50**
		1.5	18	12.84	13.50	14.67	**6.09**	2.66	2.79	2.94	**1.69**
			19	12.98	13.24	14.35	**6.12**	2.61	2.72	2.86	**1.66**
BBD	3	0.5	18	8.67	**1.83**	2.28	3.36	2.37	**1.07**	1.20	1.55
			19	8.17	**1.70**	2.14	3.16	2.24	**1.02**	1.15	1.48
		1	18	9.31	**3.46**	4.36	3.88	2.43	**1.47**	1.65	1.62
			19	8.60	**3.32**	4.15	3.63	2.30	**1.42**	1.60	1.56
		1.5	18	9.66	**4.97**	6.11	4.26	2.48	**1.75**	1.96	1.69
			19	8.95	**4.72**	5.86	3.98	2.34	**1.68**	1.90	1.63
H416A	4	0.5	18	9.50	**6.57**	7.03	13.04	2.11	**1.87**	1.96	2.77
			19	9.41	**6.41**	6.88	15.12	2.11	**1.82**	1.92	2.97
		1	18	9.92	**12.58**	13.45	13.17	2.19	**2.58**	2.71	2.76
			19	10.09	**11.94**	12.74	15.40	2.22	**2.48**	2.61	2.95
		1.5	18	10.15	**18.39**	19.50	13.25	2.23	**3.12**	3.26	2.74
			19	9.76	**17.40**	18.46	14.64	2.21	**3.00**	3.13	2.86
H416B	4	0.5	18	**10.78**	346.95	347.20	16.52	**2.23**	9.99	10.05	3.05
			19	**10.36**	281.69	281.97	18.88	**2.22**	8.79	8.85	3.29
		1	18	**11.26**	578.62	579.40	16.62	**2.32**	12.88	12.95	3.01
			19	**10.82**	498.91	500.01	18.88	**2.30**	11.64	11.73	3.22
		1.5	18	**11.13**	816.31	817.97	15.99	**2.33**	15.23	15.33	2.93
			19	**10.56**	691.47	692.22	17.69	**2.30**	13.88	13.97	3.08
H416C	4	0.5	18	**10.90**	425.14	751.42	16.59	**2.25**	9.62	11.83	3.06
			19	**10.33**	3443.86	3957.55	18.79	**2.22**	25.64	26.30	3.30
		1	18	**10.63**	435.54	770.16	15.85	**2.27**	9.95	12.10	2.95
			19	**11.46**	3392.17	3874.29	19.43	**2.35**	24.96	25.48	3.25
		1.5	18	**11.54**	485.31	819.70	16.40	**2.37**	10.55	12.60	2.96
			19	**10.70**	3932.13	4604.19	17.87	**2.32**	26.34	27.34	3.12



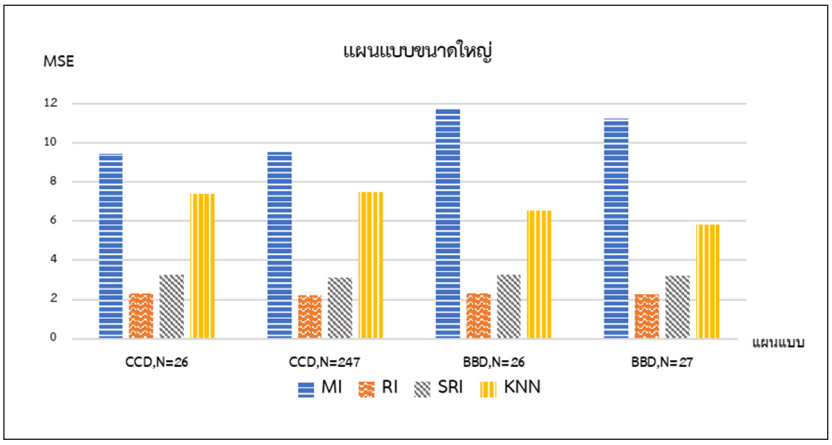
รูปที่ 2 ค่า MSE ของวิธีประมาณค่าสูญหายทั้ง 4 วิธี ในแผนแบบพื้นผิวตอบสนองขนาดกลาง ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1

ตารางที่ 3 ค่า MSE และ MAE ในแผนแบบพื้นผิวตอบสนองขนาดใหญ่ ($N = 26, 27$) เมื่อ $k = 4$

แผนแบบ	k	σ^2	N	เกณฑ์ที่ใช้เปรียบเทียบ							
				MSE				MAE			
				MI	RI	SRI	KNN	MI	RI	SRI	KNN
CCD	4	0.5	26	8.91	**1.16**	1.67	6.89	2.37	**0.85**	1.03	2.05
			27	8.74	**1.11**	1.60	6.77	2.32	**0.84**	1.01	2.04
		1	26	9.46	**2.30**	3.27	7.40	2.40	**1.20**	1.43	2.12
			27	9.54	**2.23**	3.12	7.50	2.37	**1.18**	1.40	2.12
		1.5	26	10.25	**3.29**	4.72	8.07	2.47	**1.44**	1.72	2.21
			27	9.66	**3.22**	4.57	7.72	2.38	**1.42**	1.70	2.16
BBD	4	0.5	26	11.03	**1.21**	1.69	6.07	2.67	**0.87**	1.03	2.09
			27	10.58	**1.17**	1.66	5.40	2.59	**0.86**	1.02	1.94
		1	26	11.71	**2.32**	3.25	6.56	2.75	**1.20**	1.43	2.15
			27	11.24	**2.25**	3.19	5.81	2.66	**1.18**	1.41	1.99
		1.5	26	12.14	**3.46**	4.89	6.93	2.81	**1.47**	1.76	2.20
			27	11.59	**3.30**	4.74	6.25	2.71	**1.44**	1.73	2.05

4. อภิปรายผลและสรุป

ในการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าสูญหายทั้ง 4 วิธีในแผนแบบพื้นผิวตอบสนองแบบ CCD, SCD, BBD และ Hybrid ที่มีจำนวนปัจจัยเท่ากับ 3 และ 4 ปัจจัย ในขอบเขตทรงกลม โดยพิจารณาจากค่า MSE และ MAE พบว่าทั้งสองเกณฑ์ให้ผลการศึกษาที่สอดคล้องกัน โดยในแผนแบบพื้นผิวตอบสนองขนาดเล็ก ($12 \le N \le 14$) พบว่า ในแผนแบบ SCD วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายดีที่สุด แผนแบบ Hybrid วิธี MI มีแนวโน้มที่จะประมาณค่าสูญหายได้ดียกเว้นแผนแบบ H311A และ H311B ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 0.5 ส่วนในแผนแบบพื้นผิวตอบสนองขนาดกลาง ($16 \le N \le 19$) พบว่า ใน



รูปที่ 3 ค่า MSE ของวิธีประมาณค่าสูญหายทั้ง 4 วิธี ในแผนแบบพื้นผิวตอบสนองขนาดใหญ่ ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1

แผนแบบ CCD และ BBD วิธี RI มีแนวโน้มที่จะประมาณค่าสูญหายได้ดี ยกเว้นแผนแบบ BBD ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 1.5 สำหรับแผนแบบ SCD วิธี KNN มีแนวโน้มที่จะประมาณค่าสูญหายได้ดี ซึ่งสอดคล้องกับงานวิจัยของอุษณีย์ [10] ที่ได้ศึกษาเปรียบเทียบวิธีการประมาณค่าสูญหายแบบนอนอิกนอร์เรเบิลในการวิเคราะห์การถดถอยเชิงเส้นพบว่า วิธี KNN มีประสิทธิภาพในการประมาณค่าสูญหายที่ดีขึ้นหากส่วนเบี่ยงเบนมาตราฐานของความคลาดเคลื่อนมีค่าสูงขึ้น ส่วนในแผนแบบ Hybrid วิธี MI มีแนวโน้มที่จะประมาณค่าสูญหายได้ดี ยกเว้นแผนแบบ H416A ที่มีความแปรปรวนของความคลาดเคลื่อนเท่ากับ 0.5 และในแผนแบบพื้นผิวตอบสนองขนาดใหญ่ ($N \geq 26, 27$) พบว่า ทั้งในแผนแบบ CCD และ BBD วิธี RI มีประสิทธิภาพในการประมาณค่าสูญหายได้ดีที่สุด นอกจากนี้ยังพบว่า วิธีการประมาณค่าสูญหายทั้ง 4 วิธี จะมีประสิทธิภาพสูงขึ้นถ้ามีการทำซ้ำเพิ่มมากขึ้นที่จุดศูนย์กลางของแผนแบบ (n_c) และมีประสิทธิภาพในการประมาณค่าสูญหายลดลงหากมีความแปรปรวนของความคลาดเคลื่อนเพิ่มขึ้น และมีจำนวนปัจจัยเพิ่มขึ้น ซึ่งสอดคล้องกับงานวิจัยของ Yakubu [11] ที่ได้ศึกษาผลกระทบของการมีข้อมูลสูญหายที่ส่งผลต่อความสามารถในการประมาณค่าผลตอบสนองในแผนแบบ Central Composite Design

5. กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณโครงการพัฒนากำลังคนด้านวิทยาศาสตร์ (ทุนเรียนดีวิทยาศาสตร์แห่งประเทศไทย) และขอขอบคุณอาจารย์และเจ้าหน้าที่ในภาควิชาสถิติคณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

เอกสารอ้างอิง

[1] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response Surface Methodology : Process and Product Optimization Using Designed Experiments*, 3rd ed. New York, 2009.

[2] H. O. Hartley, “Smallest composite design for quadratic response surfaces,” *Biometrics*, vol. 15, no. 4, pp. 159–171, 1959.

[3] B. Chomtee, “Comparison of design optimality criteria of reduced models for response surface designs in a spherical design region,” Ph.D. dissertation, Montana State University, 2003.

[4] W. Pardubsri, “A Comparison study of spherical response surface designs for a set of reduced models by graphing methods,” M.S. thesis,

Department of Statistics, Faculty of Sciences Kasetsart University, 2015 (in Thai).

[5] B. W. Bolch and C. J. Huang, *Multivariate Statistical Method for Business and Economics*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1974, p. 329.

[6] R. J. A. Little and B. D. Rubin, *Statistical Analysis with Missing Data*, John Wiley & Sons, Inc., New York, 2nd ed. 2002, pp. 60–61.

[7] R. Lumjaisue,"Comparison of missing data estimation methods for the multiple regression analysis with missing at random dependent variable," *Thammasat International Journal of Science and Technology*, vol. 25, no. 5, pp. 676–777, 2017 (in Thai).

[8] W. Chaimongkol, "Three composite imputation methods for item nonresponse estimation in sample surveys," Ph.D. dissertation, Department of Applied Statistics, National Institute of Development Administration, Bangkok, 2005 (in Thai).

[9] P. Jonsson and C. Wohlin, "An evaluation of k-nearest neighbor imputation using likert data," in *Proceedings of the 10th International Symposium on Solfware Metrics*, 2004, pp. 1530–1435.

[10] A. Wongarmart, "Comparison of the estimation methods for nonignorable missing data in multiple linear regression," M.S. thesis, Department of Statistics, Faculty of commerce and accountancy Chulalongkorn University, Bangkok, 2012 (in Thai).

[11] Y. Yakubu, A. U. Chukwu, B. T. Adebayo, and A. G. Nwanzo, "Effects of missing observations on predictive capability of central composite designs," *International Journal on Computational Science & Applications*, vol. 4, no. 6, pp. 1–18, 2014.