Metodología robusta para superficies de respuestas

MC Manuel R. Piña[1], Dr. Manuel A. Rodríguez[2] y Dra. Elsa M. Benavides[2]

RESUMEN

Esta investigación enfatiza el uso del diseño central compuesto y la regresión ridge para ajustar un modelo polinomial de segundo orden para representar la superficie de respuesta. Esto corresponde a la necesidad de una metodología estandarizada para mejorar la Metodología de Superficies de Respuesta (RSM). La RSM es útil en la determinación del mejor nivel de factores para determinar la salida de la variable de respuesta. Nosotros describimos un conjunto de doce pasos básicos para la aplicación efectiva de la metodología de superficies de respuesta. Una discusión de cada uno es incluido y su alcance también.

Palabras Clave: Metodología de Superficies de Respuestas, Regresión Ridge, Mínimos Cuadrados, Multicolinealidad, Diseño Central Compuesto.[12]

1. INTRODUCCIÓN

La Metodología de Superficies de Respuestas (MSR), es un conjunto de herramientas estadísticas y matemáticas, utilizadas para optimizar una variable de respuesta sujeta a varias variables predictoras. La MSR, se utiliza cuando las relaciones entre las variables, no son completamente entendidas como para representarlas de manera directa a través de un modelo matemático exacto, si no que es necesario construir un modelo empírico para aproximar su comportamiento. Para aproximar este comportamiento, la MSR utiliza generalmente un diseño factorial fraccionado de resolución III para determinar la subregión de las variables predictoras para la cual la variable de respuesta presenta un óptimo. Esta subregión es alcanzada a través del ajuste de un modelo polinomial de primer orden a este diseño, ajustado por Mínimos Cuadrados (MC) y la

[1] Manuel R. Piña Monarrez, cursa el grado de Doctor en Ciencias con Especialización en Ingeniería Industrial. Este articulo se basa en su disertación doctoral "Metodología Robusta para Superficies de Respuestas". Programa Doctoral, Instituto Tecnológico de Ciudad Juárez, México. romanpin1970@yahoo.com.mx
[2] División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Ciudad Juárez, México.
mrodriguez@itcj.edu.mx, ebenavides@itcj.edu.mx

aplicación iterativa del método de ascenso acelerado (ver Box y Draper 1987). Desafortunadamente, debido a las pocas corridas experimentales que el diseño requiere y al hecho de que la falta de ajuste del modelo de primer orden en esa subregión, es determinada cuando los efectos de interacción y/o cuadráticos puros son significantes, originan un incremento en el error tipo II si ajustamos un modelo polinomial completo de segundo orden cunado solo la iteración es significante (o viceversa) (Ganzach, 1998). Dado que el modelo completo de segundo orden con que se modelan las superficies de respuestas, presenta de forma inherente el problema de multicolinealidad, por lo que el punto estacionario encontrado puede estar lejos del óptimo, el polinomio deberá de ser ajustado a través de la regresión ridge (RR). Para el ajuste del polinomio a través de RR, recomendamos utilizar un diseño central compuesto (DCC), dadas las características de ortogonalidad, optimalidad D y optimalidad G que este presenta (Myers y Montgomery, 1995). Para un análisis detallado de la MSR, referimos al lector a Box y Wilson (1951); Box, Hunter y Hunter (1978); y Box y Draper (1987) entre otros.

2. PASOS ESTÁNDAR DE LA METODOLOGÍA DE SUPERFICIES DE RESPUESTAS

Sin perdida de generalidad, permítanos asumir que el modelo a optimizar es un modelo de varias variables del tipo :

$$Y = f(\xi_1, \xi_2, ..., \xi_k) + \varepsilon_{ij} \qquad (2.1)$$

donde ξ_j , son las variables naturales del proceso. La respuesta esperada en variables codificadazas con $\mu = 0$ y $\sigma^2 I$ constante esta dada por:

$$E(Y) = f(x_1, x_2, ..., x_k) = \eta \qquad (2.2)$$

Así, la superficie de respuesta para estas variables esta dada por:

$$\eta = f(x_1, x_2, ..., x_k) \qquad (2.3)$$

donde η, es la superficie de varias variables a optimizar y $(x_1, x_2, ..., x_k)$, son las variables independientes codificadas (factores de diseño). El componente ε_{ij} dado en (2.1), representa el error aleatorio considerado $N(0, \sigma^2)$. La naturaleza secuencial de la MSR, inicia cuando existe una característica de interés a ser medida y continua con la generación de ideas para determinar el conjunto de factores significantes que determinan su comportamiento. Los factores que se creen determinan el comportamiento de la variable de respuestas, son analizados en diseños de experimentos iterativos para a través del análisis de varianza (ANOVA) determinar cuales son significantes. Myers y Montgomery (1995) a este proceso iterativo, le dan el nombre de "Paso Cero". Para fines de este articulo, permítanos asumir que el conjunto de factores considerado en el diseño experimental, son los factores significantes que determinan el comportamiento de la variable de respuesta. Por lo que los pasos estándar para la aplicación de la MSR son:

1. Inicio. Determine las variables naturales (ξ_j).- Para este propósito, la metodología seis sigma, propone el uso del despliegue de la función de calidad (QFD por sus siglas en ingles), el mapeo del proceso, el diagrama de pareto y el diagrama de ishikawua. El paso incluye:

a. Determinar los niveles de las variables naturales (ξ_j). El establecimiento de los niveles, depende de la naturaleza de la experimentación, de la región de operabilidad y de la región a ser modelada, por lo que el conocimiento del investigador del proceso bajo estudio y el alcance del objetivo que se persigue, son determinantes (Box y Draper, 1987).

b. Codifique las variables naturales. Las variables codificadas, permiten una mejor comparación entre ellas, se recomienda que la codificación entre los niveles sea equidistante. La codificación de las variables esta dada por:

$$X_i = \frac{\xi_i - \overline{\xi}}{S} \qquad (2.4)$$

donde ξ_j las variables naturales y S, es la desviación entre la variable y su media $\overline{\xi}$.

2. **Ajuste un Polinomio de Primer Grado.** Para el ajuste de este polinomio, se recomienda utilizar un diseño experimental fraccionado de resolución III ya que la varianza de la respuesta esperada para este diseño es mínima (Khuri y Cornell, 1987). Además, este diseño requiere un numero pequeño de corridas experimentales y puede ser fácilmente incrementado a un diseño central compuesto para ajustar el modelo polinomial completo de segundo orden (Myers y Montgomery, 1995). El modelo de primer orden es:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \varepsilon \qquad (2.5)$$

3. **Determine la Adecuación del Modelo de Primer Orden.** Antes de utilizar el modelo polinomial de primer orden para moverse dentro de la región de operabilidad hacia el óptimo, determine su adecuación a través del análisis de varianza. Para esto es necesario que el diseño utilizado no sea saturado y que se hayan realizado replicas genuinas (varias corridas experimentales para un nivel central de las variables) con la finalidad de determinar el error puro. Para determinar la adecuación del modelo recomendamos el procedimiento descrito por Montgomery, Peck y Vinnin (2002), en donde a través del ANOVA, se determina si el comportamiento de la respuesta en la

región actual de interés, presenta interacción y/o curvatura que refleje falta de ajuste del modelo.

i) Si el modelo no muestra falta de ajuste, pero los coeficientes estimados son cercanos a cero, es una indicación de que no hay dirección de mejoramiento en ese punto y el proceso de optimización, deberá de iniciarse una vez más en el paso 1.

ii) Si la falta de ajuste es debida solo a la interacción entre los factores y no a curvatura pura, la varianza entre los niveles de los factores deberá de ser reducida y otro diseño experimental deberá de llevarse acabo. No ajuste bajo esta situación un polinomio completo de segundo orden ya que esto incrementa el error tipo II (Ganzach, 1998).

iii) Si la inadecuación del modelo es dada porque los términos de interacción y cuadráticos puros son significantes pase al paso 6.

4. Resuelva la Inadecuación del Modelo. Si la falta de ajuste es solo significante para los terminas de interacción, la varianza de las variables naturales (S), deberá de ser reducida. Es decir se minimiza la región de exploración y de esa forma el modelo lineal proporcionara un buen ajuste.

5. Aplique el Método de Ascenso Acelerado para Acceder a las cercanías del Óptimo. Si el modelo de primer orden es aceptado para representar la superficie en la región actual de experimentación, aplique el proceso iterativo del método de ascenso acelerado a este polinomio para acceder a las cercanías del óptimo, avanzando en la dirección dada por el signo de los coeficientes de regresión (Si es descendente, deberá de tomar los signos opuestos), con incrementos proporcionales a la magnitud de los coeficientes. El nuevo punto así estimado, es tomado como el nuevo punto central del diseño en $n+1$ iteración y el procedimiento es repetido hasta que la falta de ajuste del polinomio este presente. Otra opción, consiste en definir el avance dando incrementos equivalentes a la distancia del punto central del diseño en la dirección de los signos de los coeficientes de regresión al punto de intersección de la esfera de radio 1 dada por:

$$\sum_{i=1}^{k} X^2 = 1 \qquad (2.6)$$

El procedimiento iterativo termina cuando la falta de ajuste es significativa o cuando en dos iteraciones consecutivas, el mejoramiento en la variable de respuesta es mínimo, comparado con un límite pre-establecido por el investigador. Para un detallado análisis, referimos al lector a (Box y Draper, 1987; Myers y Montgomery, 1995; Khuri y Cornell, 1987) entre otros.

i) En el caso en el que a través de la aplicación iterativa del método de ascenso acelerado, la falta de ajuste del modelo de primer orden sea significante, un diseño central compuesto deberá de ser utilizado para ajustar un modelo completo de segundo orden, determinar su adecuación a través del ANOVA.

ii) Si el modelo completo de segundo orden proporciona un buen ajuste, pero los términos de segundo orden, son pequeños comparados con los de primer orden, se deberá de aplicar el método de ascenso acelerado hasta que los términos de segundo orden sean potencialmente importantes.

6. Aproxime la Superficie a través de un Modelo Completo de Segundo Orden. Cuando el modelo de segundo orden del tipo:

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_i + \sum_{i=1}^{k} \hat{\beta}_{ii} X_i^2 + \sum_{I=1}^{k} \sum_{J=1}^{k} \hat{\beta}_{ij} X_{ij} \qquad (2.7)$$

El cual puede ser escrito en forma matricial como $\hat{Y} = \hat{\beta}_0 + X^t\hat{\beta} + X^tBX$ va ha ser ajustado para modelar la superficie, recomendamos:

a) Utilizar un Diseño Central Compuesto. Este diseño es recomendado, dadas sus propiedades de:

i) Ortogonalidad (Obtención de los coeficientes principales en forma separada).

ii) Optimalidad D (maximiza el determinante $|X^tX|$ y de esa forma, la región de confianza del elipsoide de donde se determinan los coeficientes del polinomio dada por $(\hat{\beta}-\beta)X^tX(\hat{\beta}-\beta)$).

iii) Optimalidad G (obtiene la varianza de la predicción mínima del modelo).

iv) Es de fácil construcción para la rotabilidad (varianza de la predicción constante sobre la esfera de radio 1) cuando se le agregan puntos axiales dados por:

$$\alpha = \sqrt[4]{2^k} \qquad (2.8)$$

donde 2^k es el numero de puntos factoriales del diseño. El número de puntos centrales recomendados para dar estabilidad a la varianza de la predicción es de 3 a 5 puntos.

b) Escale la matriz de regresores y la variable y, a través del método de escalonamiento unitario para evitar el deterioramiento innecesario de la ordenada al origen y minimizar el deterioramiento de la matriz de covarianzas (X^tX).

c) Ajuste un modelo polinomial de segundo orden. El ajuste de este polinomio, deberá de realizarse a través del método de regresión ridge, dado que es el mejor método de estimación sesgada (Golam-Kibria, (2003). El polinomio completo de segundo orden, presenta de forma inherente el problema de multicolinealidad, el cual causa deterioramiento en la matriz de precisión

$(X^t X)^{-1}$, generando inestabilidad en los coeficientes estimados debido a los factores de inflación de la varianza. Ver por ejemplo Hoerl y Kennard, 1970a; Hoerl y Kennard, 1970b; Obenchain, 1975; Montgomery, Peck y Vinnin, 2002; y Piña, Rodríguez y Díaz, 2005-a.

d) Determine la constante de proporcionalidad *K.* Entre los muchos métodos de estimarla (Además de la inspección directa de la traza), nosotros proponemos estimarla a través de:

$$k = \frac{p\sigma^2}{\beta^t (X^t X)\beta} \tag{2.9}$$

(ver Lawless y Wang, 1976; Piña, Rodríguez y Díaz, 2005-a). Para el analisis de otros metodos de estimar K, ver Marquardt, 1970; Obenchain, 1975; Hemmerle, 1975; Hoerl, Kennard y Baldwin, 1975; Hoerl y Kennard, 1976; Hemmerle, 1978 entre otros. En estudios recientes, Rubio y Firinguetti (2002) y Golam Kibria (2003), encontraron que el método propuesto por Lawles y Wang (1976) es mejor para determinar el valor de K que el propuesto por Hoerl y Kennard (1976) cuando la multicolinealidad presente en el diseño es pequeña, como lo es en este caso.

7. Determine la Adecuación del Modelo Polinomial de Segundo Orden. Como en el caso del modelo de primer orden, la inadecuación del modelo, es determinada a través del ANOVA.

8. Resuelva la Inadecuación del Modelo Polinomial de Segundo Orden. Si el modelo no es el correcto para representar la superficie en esa región, es necesario reducir la región de experimentación, reduciendo la varianza (S) de los factores (Neddermeijer, Oortmarssen, Piersma y Dekker, 2000).

9. Realice el Análisis Canónico de la Superficie. Una vez que el modelo completo de segundo orden es adecuado, se determina de este modelo, el polinomio canónico para la optimización de la superficie a través de explorar y determinar la localización y naturaleza del punto estacionario. La forma matricial del modelo completo de segundo orden a ajustar es:

$$\hat{Y} = \hat{\beta}_0 + X^t \hat{\beta} + X^t BX \qquad (2.10)$$

donde:

$X^t = (x_1, x_2, \ldots, x_k)$, $\hat{\beta}^t = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k)$ y

$$B = \begin{bmatrix} \beta_{11} & \beta_{12}/2 & \ldots & \beta_{1k}/2 \\ & \beta_{22} & \ldots & \beta_{2k}/2 \\ & & \ddots & \vdots \\ sym & & & \beta_{kk} \end{bmatrix}, \qquad (2.11)$$

El procedimiento para obtener el polinomio canónico del modelo dado en (2.10), consiste primero en diferenciar este polinomio con respecto a X^t y establecer la derivada igual a cero para obtener la solución del punto estacionario como:

$$x_s = -\frac{1}{2}\hat{B}^{-1}\hat{\beta} \qquad (2.12)$$

Segundo, determinar si el punto estacionario definido en (2.12) es el óptimo del sistema cuadrático. Para ello suponga que $(\lambda_1, \lambda_2, \ldots, \lambda_k)$, son los eigenvalores de la matriz B dada en (2.12), por lo que si todos son positivos, x_s es un minino, si todos son negativos, x_s es un máximo y si al menos uno es de diferente signo, x_s es un punto de silla. Para un análisis detallado de este punto ver Box y Draper (1987).

i) Si el punto estacionario x_s es un punto de silla o si cae fuera de la región de experimentación, no se deberá de tomar como el nuevo punto central para el análisis y optimización de la superficie. Para este caso, deberá de ajustarse un polinomio canónico en su forma A (polinomio rotado sin términos de productos cruzados), para determinar la nueva dirección de búsqueda del óptimo a trabes de la aplicación del método de ascenso acelerado a este polinomio.

ii) Si el punto estacionario x_s encontrado, está dentro de la región de experimentación, y no representa un punto de silla, deberá de ajustarse un polinomio canónico en su forma B (polinomio rotado y sin los términos lineales) para realizar la exploración en las fronteras de x_s, tomándolo como el punto central del diseño central compuesto realizado.

iii) Si el punto x_s es un máximo, pero cae fuera de la región de experimentación, no es recomendable extrapolar, por lo que a través de los multiplicadores de Lagrange, deberá de ajustarse una esfera de radio 1 centrada en $(0, 0,..., 0)$, las cuales son el centro del diseño, para explorar con ella en las cercanías de x_s. Así, el punto óptimo encontrado dentro de esa esfera, deberá de utilizarse como el nuevo centro de del diseño y aplicar el proceso iterativo del método de ascenso acelerado hasta que x_s caiga dentro de la región de experimentación o lleguemos hasta el limite de la región de operabilidad del sistema o proceso analizado. Una vez que x_s esté dentro de la región de experimentación, un modelo canónico en su forma B, deberá de ser ajustado para realizar la exploración del sistema. (Box y Draper, 1987).

10. Después de Aceptar el Punto Estacionario. Una vez que el punto estacionario x_s es aceptado, este deberá de tomarse como centro del diseño y deberá de realizarse la caracterización del sistema que ese punto representa, de modo tal que:

i) Si el sistema de loma encontrado es un sistema estacionario para alguno de los factores (Algunos de los eigenvalores son cero, esto verificado de sus intervalos de confianza) un polinomio canónico es ajustado sin los eigenvalores que son cero y el mejor nivel de los demás factores que optimizan la respuesta en ese punto, deberá de ser determinado. Para esta situación existen óptimos alternativos, de manera que es posible optimizar alguna otra característica secundaria del proceso que se desee. (Box y Draper, 1987).

ii) Si el sistema de loma encontrado es un sistema ascendente (o descendente), la búsqueda del óptimo, deberá de realizarse llevando a cabo nuevos experimentos en la dirección ascendente (o descendente) del sistema, hasta el limite de la región de experimentación o hasta que otra característica (como costo) de la experimentación sea infactible. El óptimo para este sistema, está en el límite de la región de experimentación, justo en la dirección del gradiente del polinomio canónico ajustado (Box y Draper 1987).

iii) Si el punto x_s es un sistema estacionario, sus coordenadas representan el mejor nivel de los factores que optimizan la respuesta.

11. Criterio de Parar. La búsqueda iterativa del óptimo, generalmente se detiene después de haber ajustado solo un modelo polinomial de segundo orden (Fu 1994). Como quiera, nosotros concordamos con Neddermeijer, Oortmarssen, Piersma y Dekker (2000) en detener la búsqueda del óptimo cuando:

i) El valor de la respuesta estimada en la iteración $n+1$ no mejore sustancialmente la respuesta.

ii) La región de interés sea demasiado pequeña.

iii) Cuando existan restricciones de capital o de proceso.

12. Determine el Nivel Óptimo de los Factores. Tome como nivel óptimo de los factores, las coordenadas de X_s encontradas en el paso 11 y determine la respuesta óptima esperada. Estos niveles de los parámetros, son los parámetros bajo los cuales el proceso deberá de estar trabajando.

3. DISCUSIÓN

Dada la naturaleza iterativa de la Metodología de Superficies de Respuestas (MSR) y a la falta de un procedimiento estandarizado para su efectiva aplicación, existe una necesidad inmediata para estandarizar este proceso. Existe una gran cantidad de investigaciones realizadas para mejorar partes del proceso de la MSR, pero poco se ha hecho para estandarizarla. Neddermeijer, Oortmarssen, Piersma y Dekker (2000) proponen un procedimiento secuencial para la aplicación de la MSR en la optimización por simulación, pero en su procedimiento, no toman en cuenta la multicolinealidad inherente que el polinomio completo de segundo orden presenta. Ellos sugieren realizar el ajuste del polinomio a través de Mínimos Cuadrados. Nuestro objetivo en este artículo consistió en establecer el ajuste del polinomio cuadrático a través del método de Regresión Ridge como un paso estándar de la MSR y utilizar el Diseño Central Compuesto para representar la región de experimentación, así como establecer un conjunto de 12 pasos básicos para la aplicación efectiva de la MSR.

4. CONCLUSIONES

En este artículo, establecemos un conjunto básico de 12 pasos para la aplicación efectiva de la Metodología de superficies de Respuestas y damos un breve análisis de cada uno. El objetivo del artículo consiste en hacer énfasis en la necesidad de asegurar que los términos cuadráticos puros, sean dominantes en la región de interés donde se encuentra el óptimo, con la finalidad de minimizar el error tipo II en la decisión. Los argumentos y referencias dadas aquí, se realizan con la finalidad de mostrar la necesidad de utilizar no solo el diseño central compuesto (DCC), si no también realizar el ajuste del polinomio completo de segundo orden a través de método de Regresión Ridge (RR), dándole así estabilidad a sus coeficientes estimados y como consecuencia confiabilidad al modelo canónico y al punto estacionario Xs que determinan las condiciones operacionales del proceso o sistema bajo estudio.

BIBLIOGRAFÍA:

Box G. E. P. and Draper N. R. 1987. Empirical Model-Building and Response-surfaces, New York: John Wiley & Sons.

Box G. E. P. and Wilson K. B. 1951. On the Experimental Attainment of optimum Conditions: Journal of the Royal Statistical Society, Series B 13(1): 1 – 38.

Box G. E. P., Hunter W. G. and Hunter J. S. 1978. Statistics for Experimenters: An introduction to design, data analysis, and model building, New York: John Wiley & Sons.

Fu M. C. 1994. Optimization via Simulation: a review. Annals of Operations Research 53: 199 – 247.

Ganzach Y. 1998. Non-linearity, Multicolineality and the Probability of the Type II Error in Detecting Interaction. Journal of Management. Vol. 24, No 5, 615-622.

Golam-Kibria B. M. 2003. Performance of Some New Ridge Regression Estimators. Communication in statistics: Vol. 32, No 2, pp 419 – 435.

Hemmerle W. J. 1975. An Explicit Solution for Generalized Ridge Regression. Technometrics, Vol. 17, No 3 pp 309–314.

Hemmerle W. J. and Brantle. 1978. Explicit an Constrained Generalized Ridge Estimation. Technometrics, Vol. 20, No 2 pp 109 – 120.

Hoerl, A. E. and R. W. Kennard. 1970a. Ridge Regression: Biased estimation for non-orthogonal problems. Technometrics, Vol. 12, No , 55-67.

Hoerl, A. E. and R. W. Kennard. 1970b. Ridge Regression: Applications to non-orthogonal problems, Technometrics, Vol. 12, No 1, 69-82.

Hoerl, A. E., R. W. Kennard and K. F. Baldwin. 1975. Ridge Regression: Some Simulations. Communication in statistics, 4(2), 105-123.

Hoerl, A. E. and R. W. Kennard. 1976. Ridge Regression Iterative Estimation of the Biased Parameter. Communication in statistics, A5(1), 77-88.

Khuri A. I. and Cornell A. 1987. Response-surfaces: Design Analyses".

Montgomery D. C., Peck E. A. y Vining G. G. 2002. Introducción al Análisis de Regresión Lineal, México: Editorial Continental. Tercera edición.

Myers R. H. and Montgomery D. C. 1995. Response-surface Methodology: process and product optimization using designed experiments, New York: John Wiley & Sons.

Obenchain R. L. 1975. Ridge Analysis Following a Preliminary Test of the Communication in statistics New York: Marcel Dekker, Inc.pp.

Lawless J. F. and P. Wang. 1976. A simulation Study of Ridge and other Regression Estimators. Communication in Statistics – Theory and Method, A5 (4), 307 – 326.

Neddermeijer H. G., Oortmarssen G. J., Piersma N and Dekker R. 2000. A Framework for Response-surface Methodology for Simulation Optimization. Proceedings of the 2000 Winter Simulations Conference J. A. Joines, R. R. Barton. K. Kang, and P. A. Fishwick, Eds.

Marquardt D. W. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimations and non Linear Estimation. Technometrics, Vol. 12, No 3, pp 591–612.
Shrunken Hypothesis. Technometrics. EUA. Vol. 17, No.4, 431-441.

Piña, Rodríguez y Díaz. 2005. Superioridad de la Regresión General Ridge sobre Mínimos Cuadrados. CULCYT//Enero-Febrero, 2005, México. Año 2, No 6. 21-26

Rubio H. and L. Firinguetti. 2002. The Distribution of Stochastic Shrinkage Parameters in Ridge Regression. Communication in Statistics – Theory and Method: Vol. 39, No 9, pp 1531–1547.