

Optimización multirespuesta para mejora de la calidad. Comparación de enfoque clásico con el enfoque bayesiano y el de redes neuronales

Juan Cevallos Ampuero¹

Recibido: 14/09/12 Aceptado: 26/02/13

RESUMEN

El trabajo tiene por objetivo revisar las metodologías existentes sobre optimización multirespuesta, integrarlas en una sola y desarrollar un nuevo algoritmo que permita superar las limitaciones existentes. Para tal efecto se revisaron las metodologías de optimización estadística mediante metodología de superficie de respuesta tradicional, con diseño robusto; seguidamente se revisó la aplicación del enfoque bayesiano a lo obtenido con la estadística tradicional; y finalmente se revisaron aplicaciones de redes neuronales artificiales a casos de optimización. Luego de realizar el análisis y discusión sobre el tema se integró las tres metodologías en una sola, habiendo desarrollado un nuevo algoritmo que permite superar las limitaciones y deficiencias de los métodos anteriores. Asimismo, se compararon los resultados obtenidos con otros métodos con los que se obtendrían con el nuevo método, siendo resultado favorable. Por tanto se ha desarrollado una metodología de optimización multirespuesta que considera relaciones lineales y no lineales, que tiene las cualidades de las metodologías de la estadística tradicional, la estadística bayesiana, y las redes neuronales artificiales.

Palabras Clave: Mejora de la calidad. Optimización Multirespuesta. Estadística Bayesiana. Redes Neuronales.

MULTIPLE RESPONSE OPTIMIZATION FOR QUALITY IMPROVEMENT. COMPARATIVE BETWEEN CLASSIC APPROACH WITH BAYESIAN APPROACH AND NEURAL NETWORKS

ABSTRACT

The paper aims to review the existing methodologies for multiresponse optimization, integrate them into one and develop a new algorithm that allows to overcome the existing limitations. For this purpose we reviewed statistical optimization methodologies using the traditional response surface methodology with robust design, then reviewed the application of the bayesian approach to that obtained with traditional statistics, and finally reviewed artificial neural network applications to cases of optimization. After performing the analysis and discussion about the three methodologies were integrated into one, having developed a new algorithm to overcome the limitations and shortcomings of the previous methods. Also, we compared the results obtained with other methods with those obtained with the new method, with favorable outcome. Thus we have developed a multi-response optimization methodology that considers linear and non-linear relationships, which has the qualities of traditional statistical methodologies, bayesian statistics, and artificial neural networks.

Keywords: Quality improvement. Multiple Response Optimization. Bayesian Statistics. Neural Networks.

I. INTRODUCCIÓN

La competitividad del mundo globalizado exige la optimización de los procesos con nuevos algoritmos, y ello justamente se busca conseguir mediante el presente estudio. En la actualidad existen diversas propuestas sobre la optimización multirespuesta para la mejora de la calidad; dentro de ellas destacan el enfoque estadístico clásico y más recientemente el enfoque que utiliza estadística bayesiana. Por otro lado las redes neuronales han demostrado múltiples aplicaciones en problemas de optimización, por lo que se considera que deberían poderse aplicar en la solución de este tipo de problemas. El objetivo del trabajo es revisar las metodologías existentes sobre optimización multirespuesta, e integrarlas en una sola y desarrollar un nuevo algoritmo que permita superar las limitaciones existentes. La hipótesis es que se puede desarrollar un algoritmo nuevo para un nuevo modelo de optimización de procesos con varias respuestas, con relaciones no lineales, con estadística bayesiana y redes neuronales.

Szu Hui NG. (2010), en su trabajo, "A Bayesian Model-Averaging Approach for Multiple-Response Optimization", publicado en el Journal of Quality Technology, vol 42, N°1, desarrolla una propuesta de optimización, que introduce la incertidumbre en los datos / parámetros mediante el uso de la estadística bayesiana con Modelos de Promedios Bayesianos BMA, para ello toma en cuenta la función pérdida de Taguchi (1986) modificada con los aportes Pignatiello (1993) y Vinning(1998). Vinning, G.G. (1998), en su trabajo, "A Compromise Approach to Multiresponse Optimization", publicado en el Journal of Quality Technology, vol 30, N°4, presenta la metodología de optimización multirespuesta y Pignatiello, J.J. (1993), en su trabajo, "Strategies for Robust Multiresponse quality engineering", IIE Transactions, vol25, N°3, presenta la aplicación de la función pérdida a la optimización multirespuesta en la Ingeniería de la Calidad.

Peterson, J. (2004), en su trabajo, "A posterior predictive approach to multiple response surface optimization", publicado en el Journal of Quality Technology. vol 36, N° 2, y Raftery, A, et al (2003), en su trabajo, "Bayesian Model Averaging for Linear Regression Models",

¹ Doctor en Ingeniería, docente de la Facultad de Ingeniería Industrial UNMSM.

publicado en el Journal of the American Statistical Association. vol 92, Nº 437, introducen el uso de promedios de modelos bayesianos BMA para los modelos de regresión lineal, lo cual es muy cercano a los Análisis de Varianza ANVA, pero no es lo mismo, y de allí a la optimización hay todo un trecho por recorrer; por otro lado; Ko, Y., Kim, K. y Jun, CH. (2005), en su trabajo, "A New loss function - based method for multiresponse optimization", publicado en el Journal of Quality Technology. vol 37, Nº 1, desarrollan un nuevo enfoque de la función pérdida para aplicarse en la optimización multirespuesta.

Rajagopal, R y Del Castillo, E. (2005), en su trabajo, "Model-robust process optimization using Bayesian model averaging" publicado en Technometrics. Nº 47-2, 2005, desarrollan una metodología de optimización de modelos usando BMA.

Cevallos, J., (2008), en su trabajo, "Redes neuronales Artificiales RNA aplicadas a la mejora de la calidad", publicado en Industrial Data 2008, Nº 2, desarrolla una metodología para mejora de la calidad aplicando redes neuronales pero no se aplica la estadística bayesiana, que es justamente lo que se buscará desarrollar en el presente trabajo.

II. METODOLOGÍA

1. Revisar las metodologías existentes,
2. Comparar las metodologías clásica, bayesiana y de redes neuronales.

3. Analizar la complementación de las metodologías
4. Sistematizar un nuevo modelo

III. RESULTADOS

3.1. Aplicación del Enfoque clásico

Para este enfoque, se aplicó y probó varios casos de diversos textos que van desde superficie de respuesta única, del texto de Khuri, A.I. y Cornell, J.A., 1996; superficie de respuesta múltiple del texto de Khuri, A.I. y Cornell, J.A., 1996 y de superficie de respuesta múltiple del texto de Del Castillo, 2007. El caso reportado en el texto de Khuri, A.I. y Cornell, J.A., Reponse Surfaces, 1996, de respuesta única sobre modelado de la cantidad de alimento ingerido por ratas de laboratorio, que indica que a 12 ratas de laboratorio de tamaño y edad uniforme se les privó de alimento, excepto una hora al día, por 10 días. En el día 11, cada rata fue inoculada con una dosis de droga reductora del hambre (dosis de 0.3 ó 0.7 mg/kg) y después de un espacio de tiempo específico (tiempos de 1, 5 ó 9 horas) las ratas fueron alimentadas. Cada una de las 6 combinaciones (dosis x tiempo) fueron aplicadas a 2 ratas. El peso en gramos del alimento ingerido por cada rata fue medido. El objetivo del experimento era determinar si el cambio en la dosis de droga así como el intervalo de tiempo entre inoculaciones y alimentación tenían algún efecto sobre las ratas en términos de la cantidad de alimento ingerido. Los resultados obtenidos se muestran en la tabla 1.

Tabla 1. Pesos de alimentos ingeridos en el experimento de alimentación de ratas

Dosis de droga (mg/kg)	Intervalos de tiempo entre inoculación y alimentación (horas)			Totales
	1	5	9	
0.3	5.63 6.42	11.57 12.16	12.68 13.31	61.77
0.7	1.38 1.94	5.72 4.69	8.28 7.73	29.74
Totales	15.37	34.14	42.00	91.51

Con base a los datos obtenidos, se propuso el modelo:

$$Y_u = \beta_0 + \beta_1 X_{u1} + \beta_2 X_{u2} + \beta_3 X_{u1} X_{u2} + \beta_4 X_{u2}^2 + \epsilon_u \dots \dots u = 1, 2, \dots, 12 \tag{ec.1}$$

En notación matricial: $y = x\beta + \epsilon$ ec.2

Las ecuaciones normales fueron dadas por: $X' bX = X' Y$ ec.3

La solución las ecuaciones normales fueron: $b = (X' X)^{-1} X' Y$ ec.4

El modelo ajustado de segundo orden obtenido fue:

$$Y(x) = 8.428 - 21.369 X_1 + 1.87 X_2 - 0.195 X_1 X_2 - 0.085 X_2^2 \tag{ec.5}$$

Este modelo se puede resolver para encontrar un punto estacionario, mediante fsolve de MATLAB se obtiene $x(1)=0.7$; y $X(2)=1$.

El ratio F para evaluar la hipótesis $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ sobre el ajuste de la curva se obtuvo: $CM \text{ Regresión} / CM \text{ Residual} = 44.823 / 0.592 = 75.73$ que es mayor que el F tabular $F_{0.01,4,7} = 7.85$; por tanto se rechaza la H_0 . Luego al menos un β es distinto de 0.

Asimismo, se calculó la prueba de falta de ajuste del modelo ajustado usando las réplicas de las observaciones, usando la relación de $F = CM \text{ Falta de Ajuste} / CM \text{ Error Puro}$; obteniéndose $2.620 / (1.523/6) = 10.32$; que es mayor que el F tabular $F_{0.05,1,6} = 5.99$; por tanto se rechaza la H_0 ; es

decir, hay falta de ajuste a un nivel de significancia de 0.05.

Seguidamente se ha considerado el caso reportado en el texto de Khuri, A.I. y Cornell, J.A., *Reponse Surfaces*, 1996, de respuesta múltiple sobre concentrados de una proteína de suero. El experimento se desarrolló para investigar los efectos de la temperatura de calentamiento (x_1), nivel de pH(x_2), potencial redox (x_3), oxalato de sodio (x_4) y sulfato láurico de sodio (x_5) sobre las propiedades de formación de espuma de los concentrados de proteína de suero. Se midieron tres respuestas, el tiempo de batido (y_1), la máxima cobertura (y_2) y el porcentaje de proteína soluble (y_3). Los niveles de los valores originales y codificados se presentan en la tabla 3.

Tabla 3. Niveles originales y codificados de las variables de entrada.

Variable		Niveles codificados				
		-2	-1	0	1	2
T° calentamiento	X1	65.0	70.0	75.0	80.0	85.0
pH	X2	4.0	5.0	6.0	7.0	8.0
Potencial Redox volt	X3	-0.025	0.075	0.175	0.275	0.375
Oxalato de sodio mol	X4	0.0	0.0125	0.025	0.0375	0.05
Sulfato láurico de sodio %	X5	0.0	0.05	0.10	0.15	0.20

El diseño utilizado fue un diseño compuesto central, factorial fraccionado $\frac{1}{2}$ del factorial completo 25, con 10 puntos axiales y 5 réplicas del punto central.

La tabla 4, contiene el diseño experimental y los datos de respuestas múltiples. Se supone que los factores x_1 y x_3 son factores ruido.

Tabla 4. Diseño experimental codificado y datos de respuestas múltiples.

Factores					Respuestas		
X1	X2	X3	X4	X5	Y1	Y2	Y3
-1	-1	-1	-1	1	4.75	1082	81.4
1	-1	-1	-1	-1	4.00	824	69.6
-1	1	-1	-1	-1	5.00	953	105.0
1	1	-1	-1	1	9.50	759	81.2
-1	-1	1	-1	-1	4.00	1163	80.8
1	-1	1	-1	1	5.00	839	76.3
-1	1	1	-1	1	3.00	1343	103.0
1	1	1	-1	-1	7.00	736	76.9
-1	-1	-1	1	-1	5.25	1027	87.2
1	-1	-1	1	1	5.00	836	74.0
-1	1	-1	1	1	3.00	1272	98.5
1	1	-1	1	-1	6.50	825	94.1
-1	-1	1	1	1	3.25	1363	95.9
1	-1	1	1	-1	5.00	855	76.8
-1	1	1	1	-1	2.75	1284	100.0
1	1	1	1	1	5.00	851	104.0
-2	0	0	0	0	3.75	1283	100.0
2	0	0	0	0	11.00	651	50.5
0	-2	0	0	0	4.50	1217	71.2
0	2	0	0	0	4.00	982	101.0
0	0	-2	0	0	5.00	884	85.8
0	0	2	0	0	3.75	1147	103.0
0	0	0	-2	0	3.75	1081	104.0
0	0	0	2	0	4.75	1036	89.4
0	0	0	0	-2	4.00	1213	105.0
0	0	0	0	2	3.50	1103	113.0
0	0	0	0	0	3.50	1179	104.0
0	0	0	0	0	3.50	1183	107.0
0	0	0	0	0	4.00	1120	104.0
0	0	0	0	0	3.50	1180	101.0
0	0	0	0	0	3.00	1195	103.0

La solución que dan Khuri y Cornell, 1996, es utilizar el modelo cuadrático ajustado, para variables codificadas, para cada una de las 3 respuestas, de la forma:

$$y = \beta_0 + \sum_{i=1}^s \beta_i x_i + \sum_{i=1}^s \beta_{ii} x_i^2 + \sum_{i=1}^{s-1} \sum_{\substack{j=2 \\ i < j}}^s \beta_{ij} X_i X_j + \varepsilon$$

ec.6

Se calcula lo referente al ajuste de las variables de respuesta; primer paso del análisis multirespuesta; así se aplica la formulación multivariable, cuya ecuación se puede escribir así:

$$Y=W\Gamma+e$$

ec.7

Donde Y=[Y1:Y2:....:Yr]; donde r es el número de respuestas.

$$W=[Z1:Z2:....:Zr]$$

ec.8

$$\Gamma = \text{es la diag}(\beta_1, \beta_2, \dots, \beta_r),$$

ec.9

$$.e = [\varepsilon_1: \varepsilon_2: \dots: \varepsilon_r]$$

$$Y_i = Z_i \beta_i + \varepsilon_i; \text{ donde } i=1,2,\dots,r$$

ec.10

$$Z_i = \text{es una matriz } N \times p_i \text{ de rango } p_i; N = n^\circ \text{ corridas; } p = n^\circ \text{ chef } \beta$$

ec.11

Las filas de e son independientemente distribuidas como N(0,Σ); donde Σ es la matriz de varianza – covarianza para las r respuestas. Es una matriz rxr, cuyos (i,j)th elementos son σij (i,j=1,2,...,r)

En este caso, la matriz W, de acuerdo a la ecuación 8, y a los datos es una matriz de orden 31x63. W=[Z1:Z2:Z3] con Z1=Z2=Z3.

Para el cálculo de la falta de ajuste se aplicó el estadístico:

$$F(c) = (\text{Grados de libertad Error puro} \times \text{Suma de Cuadrados Falta de Ajuste}(c)) / (\text{Grados de libertad Falta de ajuste} \times \text{Suma de Cuadrados Error puro}(c))$$

Donde:

$$\text{Suma de Cuadrados Error puro}(c) = c'Y'KYc;$$

ec.12

$$\text{Suma de Cuadrados Falta de Ajuste}(c) = c'Y'[IN-W(W'W)-W'-K]Yc$$

ec.13

La matriz K= diag (K1,K2,...,Kn,0) y la matriz

$$K_i = |v_i + (1/v_i)J_{v_i}; i=1,2,\dots,n$$

ec.14

.c = (c1,c2,...,cr)' es un vector no cero rx1; que permite obtener:

$$Y_c = WYc + \varepsilon c$$

ec.15

.tal que, Yc= Yc, donde Yc es un vector de N observaciones de respuesta univariada

$$Y_c = \sum_{i=1}^r c_i Y_i;$$

$$.Y_c = \Gamma c; \text{ y } \varepsilon c = \varepsilon c$$

La ecuación de F(c) se puede simplificar reemplazando:

$$G_1 = Y'[IN-W(W'W)-W'-K]Y$$

$$G_2 = Y'KY$$

Luego,

$$F(c) = (\text{Grados de libertad Error puro} \times c'G_1c) / (\text{Grados de libertad Falta de ajuste} \times c'G_2c)$$

ec.16

La falta de ajuste se calcula mediante la prueba de la Raíz más larga de Roy, que se presenta en la ecuación:

$$e_{\max} (G_2^{-1}G_1) \geq \lambda_\alpha$$

ec.17

Para nuestro caso la matriz K=diag(K1,0) donde K1=I5-(1/5)J5; y 0 es una matriz cero de orden 26x26. Los

Grados de libertad del error puro son 4 y los Grados de libertad de la falta de ajuste son 6.

El valor de la Raíz más larga de Roy da 245.518. El valor crítico λ_α para esta prueba se puede obtener de la tabla de distribución beta generalizada dad en Foster (1957). Esta tabla da valores de X_α , el punto superior al $100\alpha\%$ de la distribución del eigenvalor más grande de la matriz $(G_1+G_2)-1G_1$ cuando el modelo de la ec. 2 es correcto. La relación entre λ_α y X_α está dada por:

$$\lambda = \frac{X_\alpha}{1 - x_\alpha}$$

Para el $\alpha=0.10$ de nivel de significancia, $X_\alpha=0.9884$, por lo tanto $\lambda_\alpha=85.21$. Luego, se detecta falta de ajuste a un nivel de 10%. ($245.518 > 85.21$)

Con el objeto de evaluar la contribución de las varias respuestas a la falta de ajuste se aplica la ecuación:

$$Y_{c^*} = \sum_{i=1}^r c_i^* Y_i \tag{ec.18}$$

. c^* es un eigenvalor de G_2-1G_1 , correspondiente al eigenvalor más grande: $e_{max}(G_2^{-1}G_1) \geq \lambda_\alpha$

Donde las respuestas involucradas en esta combinación lineal se consideran tienen alguna influencia sobre la falta de ajuste. Como las respuestas pueden ser medidas en diferentes unidades, la combinación lineal se puede expresar en términos variables de respuesta estandarizadas como:

$$Y_{c^*} = \sum_{i=1}^r d_i^* z_i \tag{ec.19}$$

Donde $z_i = Y_i / \|Y_i\|$; con $\|Y_i\|$ la Norma Euclidea de $(Y_i'Y_i)^{1/2}$ del vector Y_i de observaciones sobre la respuesta i th, y $d_i^* = c_i^* \|Y_i\|$ ($i=1,2,\dots,r$). El tamaño de cada coeficiente d_i^* determina la contribución de la respuesta a la falta de ajuste.

Para nuestro caso:

El eigenvector c^* correspondiente al eigenvalor $e_{max}(G_2^{-1}G_1) = 245.18$ es:

$$C^* = (3.2659, 0.0385, -0.0904)'$$

Las normas Euclideas de los vectores respuestas Y_1, Y_2 e Y_3 son $\|Y_1\|=27.60, \|Y_2\|=5929.27$ e $\|Y_3\|=517.49$. Así en términos de variables respuesta estandarizadas z_1, z_2 y z_3 la combinación lineal aplicando la ec. 19 da:

$$Y_i^* = 90.139z_1 + 228.277z_2 - 46.781z_3 \sim 0.395z_1 + z_2 - 0.205z_3$$

El símbolo \sim indica que las 2 combinaciones son proporcionales. De lo anterior se aprecia que Y_1 e Y_2 son los principales contribuyentes de la falta de ajuste. Siendo Y_2 el más influyente. Valores de $e_{max}(G_2^{-1}G_1)$ fueron calculados para todos los subconjuntos no vacíos de las 3 respuestas. Los resultados se dan en la tabla 5.

Tabla 5. Valores de $e_{max}(G_2^{-1}G_1)$ para todos los subconjuntos vacíos de las 3 respuestas.

Subconjunto	$e_{max}(G_2^{-1}G_1)$	Valores críticos $\lambda_{0.10}$
Y1,Y2,Y3	245.518*	85.21
Y1,Y2	214.307*	85.21
Y1,Y3	45.532	85.21
Y2,Y3	32.107	85.21
Y1	14.936	85.21
Y2	19.573	85.21
Y3	28.495	85.21

De lo anterior se ve que el subconjunto con 3 respuestas y el [Y1,Y2] producen significativas faltas de ajuste. Ninguna de las tres respuestas de manera individual ni los pares [Y1,Y3] e [Y2,Y3] parecen contribuir mucho a la falta de ajuste. Por tanto, en la solución deben estar presentes las 3 variables de respuesta (Y1,Y2 e Y3). Sin embargo, este ejemplo no es optimizado por Khuri y Cornell; que posteriormente recomiendan para optimizar el enfoque de la "distancia generalizada".

Este mismo caso de Khuri y Cornell, 1996, es reportado en el texto "Process Optimization", 2007, de Del Castillo, llegándose a la siguiente solución:

La metodología que propone es la del Enfoque de Respuesta Dual. Para lo cual define un modelo de regresión con q respuestas, k factores controlables y r variables de ruido, X1 y X3 considera factores ruido, de la siguiente forma:

$$Y = \theta' x^{(m)} + \Delta' z^{(m)} + \epsilon \tag{ec.20}$$

Donde θ es una matriz $p \times q$ de coeficientes para los factores controlables (cada columna contiene todos los coeficientes para una respuesta) y $x^{(m)}$ es un vector $p \times 1$ que contiene los regresores para los factores controlables en forma de modelo (hay $p = (k+1)(k+2)/2$ de tales factores para un modelo cuadrático completo (full); establece (m) para x en formato de modelo):

$$x^{(m)} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_k \\ x_1 x_2 \\ x_1 x_3 \\ \cdot \\ \cdot \\ x_{k-1} x_k \\ x_1^2 \\ x_2^2 \\ \cdot \\ \cdot \\ x_k^2 \end{bmatrix} \tag{ec.21}$$

La matriz Δ es $r(k+1) \times q$ y contiene los coeficientes de primer orden para factores ruido así como coeficientes para las interacciones de factores ruido controlable. Estos términos están combinados en una matriz única para evitar obtener un término de covarianza cruzado cuando se aplica el operador varianza. La matriz es:

$$\Delta = \begin{bmatrix} \beta_{y1z1} \beta_{y2z1} \dots \beta_{yqz1} \\ \beta_{y1z1x1} \beta_{y2z1x1} \dots \beta_{yqz1x1} \\ \beta_{y1z1x2} \beta_{y2z1x2} \dots \beta_{yqz1x2} \\ \cdot \\ \cdot \\ \beta_{y1z1xk} \beta_{y2z1xk} \dots \beta_{yqz1xk} \\ \beta_{y1z2} \beta_{y2z2} \dots \beta_{yqz2} \\ \beta_{y1z2x1} \beta_{y2z2x1} \dots \beta_{yqz2x1} \\ \beta_{y1z2x2} \beta_{y2z2x2} \dots \beta_{yqz2x2} \\ \cdot \\ \cdot \\ \beta_{y1zrx1} \beta_{y2rx1} \dots \beta_{yqzrx1} \\ \beta_{y1zrx2} \beta_{y2zrx2} \dots \beta_{yqzrx2} \\ \cdot \\ \cdot \\ \beta_{y1zrxk} \beta_{y2zrxk} \dots \beta_{yqzrxk} \end{bmatrix} \tag{ec.22}$$

Donde $\beta_{yijzjxk}$ es el coeficiente para la interacción entre x_j y x_k para la i th respuesta. Los coeficientes de primer orden para los factores ruido son los β_{yizk} por lo tanto la matriz es $r(k+1) \times q$ y es compatible con la definición de $z^{(m)}$ que se presenta a continuación.

El vector $z^{(m)}$ es $r(k+1) \times 1$ y contiene los factores ruido y sus interacciones con los factores controlables:

$$z^{(m)} = \begin{bmatrix} z_1 \\ z_1 x_1 \\ z_1 x_2 \\ \cdot \\ \cdot \\ z_1 x_k \\ z_2 \\ z_2 x_1 \\ z_2 x_2 \\ \cdot \\ \cdot \\ z_2 x_k \\ \cdot \\ \cdot \\ z_r \\ z_r x_1 \\ z_r x_2 \\ \cdot \\ \cdot \\ z_r x_k \end{bmatrix} \tag{ec.23}$$

Finalmente el vector ϵ es $qx1$ y se asume con distribución normal con media 0 y covarianza la matriz Σ_ϵ , esto es:

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_q \end{bmatrix} \sim N_q(0, \Sigma_\epsilon) \quad \text{ec.24}$$

Note que la ecuación 20 asume que todas las respuestas pueden ser aproximadamente modeladas por funciones de la misma forma, esto es, por funciones que contienen los mismos parámetros significativos.

Miro et al asume que:

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix} \sim N_r(0, \Sigma_z) \quad \text{ec.25}$$

Con Σ_z conocido de los datos históricos.

Usando las asunciones previas uno puede tomar los operadores de valor esperado y varianza de la ec.20:

$$E(Y) = \theta'x(m) \quad \text{ec.26}$$

Asimismo, el operador varianza de la ec. 20:

$$\text{Var}(Y) = \Delta' \text{Cov}(z(m)) \Delta + \Sigma_\epsilon \quad \text{ec.27}$$

Donde $\text{Cov}(z(m))$ es una matriz $[(k+1)rx(k+1)r]$. La misma que puede ser escrita como:

$$\text{Cov}(z(m)) = \Sigma_z [x(l)x'(l)] \quad \text{ec.28}$$

Donde $x(l)$ se forma mediante $k+1$ primeros elementos de $x(m)$, Σ_z es la matriz de covarianza de los factores ruido y es el producto de Kronecker o directo.

Luego:

$$\text{Var}(Y) = \Delta' [\Sigma_z (x(l)x'(l))] \Delta + \Sigma_\epsilon \quad \text{ec.29}$$

En lo sucesivo $\Sigma_z [x(l)x'(l)]$ se denotará por Σ_z

Asimismo, Del castillo en el texto "Process Optimization", 2007, con relación a la estimación de la media propone:

$$\tilde{E}(Y) = \tilde{\theta}'x \quad \text{ec.30}$$

Luego:

$$E(\tilde{E}(Y)) = \theta'x = E(Y) \quad \text{ec.31}$$

Para estimar la varianza se propone:

$$\tilde{\text{Var}}(Y) = \tilde{\Delta}' \Sigma_z \tilde{\Delta} + (1 - \text{tr}(\Sigma_z^* (X' \Delta X \Delta)^{-1})) \Sigma_\epsilon \quad \text{ec.32}$$

Con relación a la solución del problema de diseño de parámetros robustos para respuesta múltiple, Del castillo, 2007, propone: minimizar $f(\text{Var}(Y))$, sujeto a: $1 \leq E(Y) \leq u$; $x \in X$; donde u y 1 son $qx1$ vectores conteniendo los límites superior e inferior para las respuestas, $f(\cdot)$ es una función escalar y $\text{Var}(Y)$ y $E(Y)$ están dados por las ec. 29 y 26, respectivamente.

Aplicando la teoría propuesta por Del Castillo, 2007, al problema antes propuesto por Khuri y Cornell, se obtuvo:

Las matrices $\hat{\Theta}$ y $\hat{\Delta}$ fueron obtenidas después de normalizar cada respuesta con la correspondiente matriz estándar de variables L2. Si se establece como cero cualquier parámetro no significativo, por lo que las matrices de parámetros estimados fueron:

$$\hat{\Theta} = \begin{bmatrix} \dots\dots\dots Y_1 \dots\dots\dots Y_2 \dots\dots\dots Y_3 \dots\dots\dots \\ \text{Int} \dots\dots 0.1692 \dots\dots 0.1812 \dots\dots 0.1813 \\ x_2 \dots\dots 0.0068 \dots\dots -0.0031 \dots\dots 0.0145 \\ x_4 \dots\dots -0.0068 \dots\dots 0.0037 \dots\dots 0.0022 \\ x_5 \dots\dots -0.0030 \dots\dots 0.0032 \dots\dots 0.0032 \\ x_2 x_4 \dots\dots -0.0181 \dots\dots 0 \dots\dots 0 \\ x_2 x_5 \dots\dots 0 \dots\dots 0 \dots\dots 0 \\ x_4 x_5 \dots\dots -0.0125 \dots\dots 0 \dots\dots 0 \\ x_2^2 \dots\dots 0 \dots\dots -0.0022 \dots\dots -0.0071 \\ x_4^2 \dots\dots 0 \dots\dots -0.0039 \dots\dots -0.0019 \\ x_5^2 \dots\dots 0 \dots\dots 0 \dots\dots 0.0040 \end{bmatrix}$$

$$\hat{\Delta} = \begin{bmatrix} \dots\dots\dots Y_1 \dots\dots\dots Y_2 \dots\dots\dots Y_3 \\ x_1 \dots\dots 0.0460 \dots\dots -0.0297 \dots\dots -0.0159 \\ x_1 x_2 \dots\dots 0.0283 \dots\dots -0.0042 \dots\dots 0 \\ x_1 x_4 \dots\dots 0 \dots\dots 0 \dots\dots 0.004 \\ x_1 x_5 \dots\dots 0.0113 \dots\dots -0.0062 \dots\dots 0 \\ x_3 \dots\dots -0.0159 \dots\dots 0.0097 \dots\dots 0.0046 \\ x_2 x_3 \dots\dots -0.0102 \dots\dots 0 \dots\dots 0 \\ x_3 x_4 \dots\dots 0 \dots\dots 0 \dots\dots 0 \\ x_3 x_5 \dots\dots -0.0091 \dots\dots 0 \dots\dots 0.0079 \end{bmatrix}$$

La matriz de covarianza normalizada para los residuales del modelo ajustado está dada por:

$$\hat{\Sigma}_e = \mathbb{0}^{-3} \begin{bmatrix} 3.2580 & \dots & -0.7132 & \dots & -1.3049 \\ -0.7132 & \dots & 0.5304 & \dots & 0.3697 \\ -1.3049 & \dots & 0.3697 & \dots & 0.6347 \end{bmatrix}$$

Se supone que la matriz de covarianza de los dos factores ruido es conocida e igual a:

$$\Sigma_z = \begin{bmatrix} 1.0 & \dots & -0.25 \\ -0.25 & \dots & 1.0 \end{bmatrix}$$

El factor de corrección de sesgo está dado por:

$$1 - tr(\sum_z^{\otimes} (X'_{\Delta} X_{\Delta})^{-1}) = \frac{1}{12} - \frac{1}{8} (x_2^2 + x_4^2 + x_5^2)$$

El estimado de la matriz de covarianza de las respuestas está dado por:

$$\overline{Var}(Y) = \hat{\Delta}' \left[\begin{bmatrix} 1.0 & \dots & -0.25 \\ -0.25 & \dots & 1.0 \end{bmatrix} \otimes (x' x) \right] \hat{\Delta} + \left(\frac{1}{12} - \frac{1}{8} (x_2^2 + x_4^2 + x_5^2) \right) \hat{\Sigma}_e =$$

$$10^{-3} \begin{bmatrix} 2.73 + 3.39x_2 + 1.63x_5 & -1.75 - 1.29x_2 - 0.83x_5 & -0.92 + 0.20x_4 - 0.48x_5 \\ +1.05x_2^2 + 1.01x_2x_5 & -0.13x_2^2 - 0.25x_2x_5 & -0.56x_2 + 0.12x_2x_4 - 0.13x_2x_5 \\ +0.26x_5^2 & -0.084x_5^2 & +0.054x_4x_5 - 0.094x_5^2 \\ -1.75 - 1.29x_2 - 0.83x_5 & 1.12 + 0.27x_2 + 0.40x_5 & 0.59 - 0.13x_4 + 0.24x_5 \\ -0.13x_5 - 0.13x_2^2 - 0.25x_2x_5 & +0.018x_2^2 + 0.052x_2x_5 & +0.07x_2 - 0.017x_2x_4 + 0.008x_2x_5 \\ -0.084x_5^2 & +0.038x_5^2 & -0.025x_4x_5 + 0.012x_5^2 \\ -0.92 + 0.20x_4 - 0.48x_5 & 0.59 - 0.13x_4 + 0.24x_5 & 0.31 - 0.14x_4 + 0.13x_5 \\ -0.56x_2 + 0.12x_2x_4 - 0.13x_2x_5 & +0.07x_2 - 0.017x_2x_4 + 0.008x_2x_5 & +0.13x_5 + 0.016x_4^2 \\ +0.054x_4x_5 - 0.094x_5^2 & -0.025x_4x_5 + 0.012x_5^2 & -0.016x_4x_5 + 0.062x_5^2 \end{bmatrix}$$

$$+ 10^{-3} \left(\frac{1}{12} - \frac{1}{8} (x_2^2 + x_4^2 + x_5^2) \right) \begin{bmatrix} 3.2580 & \dots & -0.7132 & \dots & -1.3049 \\ -0.7132 & \dots & 0.5304 & \dots & 0.3697 \\ -1.3049 & \dots & 0.3697 & \dots & 0.6347 \end{bmatrix}$$

Suponga que los límites para las respuestas esperadas son:

$$.l = [-\infty \ 800 \ 100]; \quad u = [5.0 \ 1100 \ \infty]$$

Esto es, el valor esperado de la primera respuesta está restringido a ser menor que 5.0 minutos, el valor esperado de la segunda respuesta está restringido entre 800% y 1100% y el de la tercera respuesta debería ser mayor que 100%. Suponga se elige realizar la optimización sobre una región esférica de radio 2 para los factores controlables, así se adiciona la restricción:

$$\sqrt{x_2^2 + x_4^2 + x_5^2} \leq 2$$

La tabla 6 presenta la comparación de los valores objetivos para el ejemplo de proteína de suero.

En la tabla 6 se nota que los resultados usando las trazas son sustancialmente diferentes a los obtenidos con el determinante. Los correspondientes puntos x^* está separados por una distancia de 0.91 en unidades codificadas lo cual es bastante significativo. Los valores de los objetivos son también sustancialmente diferentes. La traza de la matriz de covarianza obtenida cuando el determinante es minimizado es cerca de 58% más grande que el obtenido cuando la traza es minimizada. Similarmente, el determinante obtenido cuando la traza es minimizada es cerca de 18% más que cuando el determinante es minimizado.

Tabla 6. Comparación de objetivos escalares para el ejemplo de Proteína de suero

Fun escalar	X*=[x ₂ *x ₄ *x ₅ *]	E(Y*)	Valor de las otras funciones de óptimo				
			Var(Y)	Tr(VarY)	Var(Y1,u)	Var(Y2,u)	Var(Y3,u)
Tr(VarY)	[.17.23 -1.98]'	[4.9 1036.8 100]'	.38e-10	.0024	1.12	2.12e4	103.8
Var(Y)	[.59.97 -1.65]'	[5.0 1026.8 100]'	.32e-10	.0038	2.14	2.73e4	73.4
Var(Y1)	[.17.17 -1.98]'	[4.9 1035.9 100]'	.40e-10	.0024	1.11	2.11e4	106
Var(Y2)	[.17.12 -1.99]'	[4.9 1035.1 100]'	.41e-10	.0024	1.11	2.11e4	108
Var(Y3)	[.77.99 -1.55]'	[4.9 1021.4 100]'	.37e-10	.0046	2.69	2.96e4	72.0

Además, note que la solución obtenida usando la traza y las varianzas de la primera y segunda respuestas son significativamente cercanas una de otra. Por ello usando las trazas como objetivo no puede capturar la variabilidad total del vector de respuestas y podría ser afectado por respuestas individuales, especialmente si más de una de ellas es minimizada cerca del mismo punto. Esto subraya la importancia de considerar la matriz de covarianza completa en lugar de las varianzas individuales, y demostrar los beneficios del enfoque multivariado propuesto sobre usar en enfoque univariante q veces en paralelo.

3.2. Enfoque Bayesiano

Para este enfoque, se aplicó y probó un caso, resuelto por dos autores distintos, pero de manera

muy similar; el primero es Miró et al, 2002, y el segundo, Del Castillo, 2007.

Miro et al en su artículo "A Bayesian Approach for Multiple response Surface Optimization in the Presence of Noise Variables" (2002), considera un ejemplo de optimización Bayesiana de un proceso de Cromatografía Líquida a alto performance HPLC, crítico en la industria farmacéutica. Es de interés estudiar de dos factores controlables, temperatura y pH, y un factor de ruido el Porcentaje de Isopropyl Alcohol % IPA; y tiene cuatro respuestas: la Resolución Crítica Rs, Tiempo Total de corrida, Ratio S/N del último pico, y factor colas del pico principal. Se usó un diseño de Box-Behnken con tres puntos centrales.

Los datos se presentan en la Tabla 7.

Tabla 7. Factores y respuestas de caso HPLC

Factores Controlables			Respuestas			
%IPA	Temp	pH	Rs	Tiempo corr.	Ratio s/N	Colas
65	30	.175	2.14	22	172	.76
65	50	.175	1.73	12	311	.88
65	40	.050	1.93	16	251	.80
65	40	.300	1.95	16	241	.80
70	40	.175	2.17	14	278	.79
70	50	.050	1.97	11	371	.86
70	30	.300	2.38	19	194	.74
70	50	.300	1.98	11	360	.86
70	30	.050	2.37	18	204	.74
70	40	.175	2.20	14	280	.78
75	40	.300	2.42	13	314	.78
75	30	.175	2.61	17	223	.73
75	50	.175	2.14	10	410	.85
75	40	.050	2.41	12	324	.78
70	40	.175	2.20	14	281	.79

Modelos de superficie de respuesta de segundo orden completos se ajustaron para las tres respuestas usando versiones codificadas de factores controlables; obteniéndose resultado de R2 superiores al 99.7% para los 4 modelos. El factor ruido %IPA se asumió con distribución normal y desviación estándar de 0.1.

La solución con el método basado en una red (grid) y Meta Modelos fue como sigue:

Un diseño 112 fue llevado a cabo con los dos regresores, donde los niveles de los regresores fueron seleccionados desde: {-1.0, -0.8,-0.6,-0.4,-0.2,0.0,0.2,0.4,0.6,0.8,1.0} y se requirieron 121

evaluación de funciones. La región A fue especificada como:

$$A = \{y = [yRs \ YTime \ Ys/N \ yTail] : yRs \geq 1.8, yTime \leq 15, ys/N \geq 300, 0.75 \leq yTail \leq 0.85\}$$

Cinco mil muestras (N=5000) fueron generadas para obtener cada valor de $\hat{p}(g(x_c))$. La mayor probabilidad obtenida fue de 0.9600 para $x_1(T^\circ)=0.4$ y $x_2(pH)=-0.4$

Un modelo de regresión logística polinomial de cuarto orden ajustado para simular $\hat{p}(g(x_c))$, para ello se uso, Minitab- Logística Binaria, obteniéndose la siguiente ecuación (ec.28):

$$\begin{aligned} \text{logit}(\hat{p}(g(x_c))) &= -20.5 + 133.7x_1 - 8.82x_2 + 38.7x_1x_2 - 265.2x_1^2 + 3.38x_2^2 - 53.9x_1^2x_2 \\ &- 8.80x_1x_2^2 + 6.12x_1^2x_2^2 + 218.8x_1^3 + 0.112x_2^3 + 24.1x_1^3x_2 - 0.145x_1x_2^3 \\ &- 69.1x_1^4 - 0.271x_2^4 \end{aligned}$$

El algoritmo de la región confiable (trust) fue usado para maximizar la regresión logística obtenida. El algoritmo fue iniciado en 50 diferentes localizaciones seleccionadas al azar dentro de la región experimental definida en la Tabla 7. En factores codificados esta región de Rc={[x1 x2]: -1≤x1≤1,-1≤x2≤1}.

En las 5 optimizaciones los algoritmos convergieron sólo a 2 puntos diferentes. Estos puntos denotados por x_c^* juntos con sus valores de predicción $\hat{p}(g(x_c))$ mediante la ec.28 son reportados en la Tabla 8. Esta tabla también tiene los estimados de $\hat{p}(g(x_c))$ obtenidos de 500000 muestras de Monte Carlo (denotadas $\hat{p}(g(x_c))$).

Tabla 8. Soluciones óptimas de (x_c^*) usando el meta modelo ec.28

X_c^*	$\hat{p}(g(x_c))$	$\hat{p}(g(x_c))$	$\hat{\sigma}_{\hat{p}(g(x_c))}$
[0.4822, 1]'	0.9684	0.9622	2.70E-04
[0.3752, -1]'	0.9769	0.9681	2.49E-04

Ambas soluciones dadas por el meta modelo, ec. 28, dan valores parecidos en términos de probabilidad. De las gráficas de las repuestas de predicción, que se muestran a continuación, se obtiene que sólo la T° tiene un efecto significativo, mientras que el pH no. Como se puede ver hay una cresta sobre la función, a lo largo del factor controlable $x_2(pH)$, lo

cual indica que solo la temperatura tiene un efecto significativo sobre el desempeño del proceso.

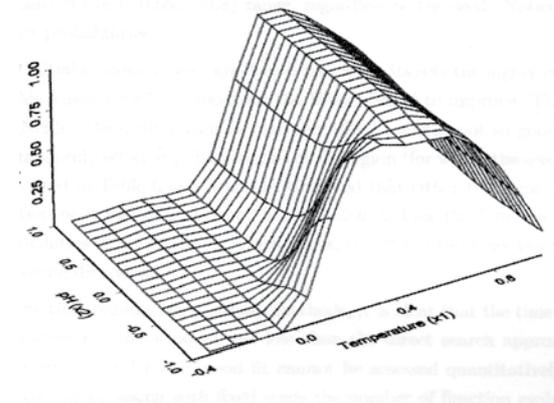
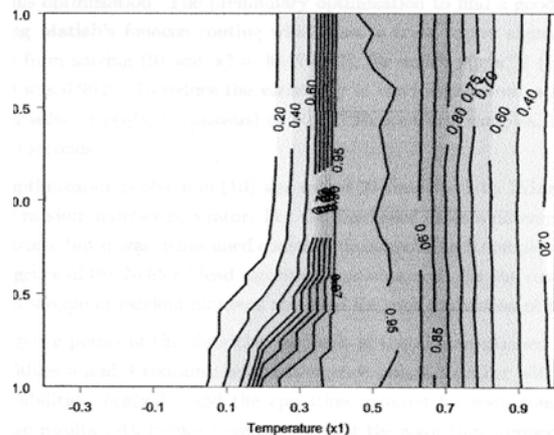


Figure 1: Plots of $\hat{p}(g(x_c))$

La optimización bayesiana usando el enfoque de los meta modelos por tanto los resultados en $w^*=[Temp^*,pH^*] = [0.482,1.0]$ para $\max P(Y \in A | data, w^*) = 0.9622$

Es de destacar que el mismo caso fue presentado por Del Castillo en su texto "Process Optimization", 2007; obteniéndose los mismos resultados.

3.3. Enfoque de Redes Neuronales Artificiales RNA

Para este enfoque, se aplicó y probó dos casos, el primero del trabajo de Lee-Ing Tong y kun-Lin Hsieh, 2000, y el segundo, se aplicó la metodología del anterior autor al caso de enfoque bayesiano HPLC del artículo de Miró et al ,2002.

Para este caso de aplicación de redes neuronales se trabajó con la metodología sugerida por Lee-Ing Tong y kun-Lin Hsieh, en su artículo "A Novel mean of Applying Neural Networks to Optimize the Multireponse Problem", publicado en el Quality Engineering, Vol.13 (2000), N° 1; que utiliza los datos de Del Castillo et al.,1996, presentan un ejemplo aplicativo sobre una faja transportadora metálica para calentamiento y adhesión de conductores de circuitos integrados en placas de plástico. El calentamiento se da en un proceso continuo a través de una corriente de aire caliente(x1), de la faja metálica a la temperatura (x2), y de un bloque de calentamiento a alta temperatura (x3). Se busca obtener la placa (de plástico) con los circuitos adheridos, pero las placas deben tener determinadas temperaturas: Si se sobre pasa demasiado dichas temperaturas: valores objetivos, las placas se pueden dañar. Estas condiciones de temperaturas que deben alcanzar las placas en diversos puntos (valores objetivo) son: Y1,Y2,Y3,Y4,Y5,Y6. Los valores objetivo utilizados son: Y1 = 190, Y2 = 185, Y3 = 185, Y4 = 190, Y5 = 185, Y6 = 185. Los datos corresponden a un diseño de experimento que es el modelo desarrollado por Box – Behnken; para 3 factores y 2 niveles para cada factor.

X1	X2	X3	Y1	Y2	Y3	Y4	Y5	Y6
40	200	250	139	103	110	110	113	126
120	200	250	140	125	126	117	114	131
40	450	250	184	151	133	147	140	147
120	450	250	210	176	169	199	169	171
40	325	150	182	130	122	134	118	115
120	325	150	170	130	122	134	118	115
40	325	350	175	151	153	143	146	164
120	325	350	180	152	154	152	150	171

80	200	150	132	108	103	111	101	101
80	450	150	206	143	138	176	141	135
80	200	350	183	141	157	131	139	160
80	450	350	181	180	184	192	175	190
80	325	250	172	135	133	155	138	145
80	325	250	190	149	145	161	141	149
80	325	250	180	141	139	158	140	148

En el presente trabajo se probó con la información proporcionada las siguientes redes con BackPropagation y con los siguientes algoritmos de entrenamiento: Basic gradient descent traingd, Basic gradient wit momentum traingdm, Adaptive rate traingdx, Resilient BP trainrp, Fletcher – Reeves conjugate gradient algorithm, traingcgf, Powell – Beale conjugate gradient algorithm traingcb, Polak - Ribiere conjugate gradient algorithm traingcp, Scaled conjugate gradient algorithm traingscg, BFGS quasi – Newton meted trainbfg, One sep secant method trainoss, Leven – Marquart algorithm trainlm, Bayesian regularization trainbr y newgrnn. Los menores MSE fueron el obtenidos con BP traingdm. La arquitectura que dio los mejores resultados fue similar a la recomendada por Tong y Hsieh: arquitectura 6-5-3 para la Red Inversa y 3-5-6 para la Red Directa.

Los resultados obtenidos son:

Aplicando la Red Neuronal BP traingdm se obtuvo X1 = 74.7560 Ft3/min, X2 = 482.9671° C, X3 = 322.0778 °C; y con ellos para los valores objetivo los siguientes: Y1 = 188.3414 °C, Y2 = 181.7106 °C, Y3 = 184.8623 °C, Y4 = 195.5162 °C, Y5 = 175.3033 °C, Y6 = 190.8334 °C.

Para los datos del caso HPLC sobre el Enfoque bayesiano del artículo de Miró et al ,2002, y aplicando el método propuesto por Lee-Ing Tong y kun-Lin Hsieh, 2001, aplicando Perceptrón Multicapa con Backpropagation y algoritmo de entrenamiento de Levenberg – Marquart ; se obtuvo como valores de los parámetros de entrada: 68.7557 para % IPA; 50.5112 °C y -0.0095 para pH; y la Red obtenida permitió predecir los valores de las respuestas, obteniéndose valores de Rs=1.8823, YTime=1.06116, Ys/N=366.2009, e Ytail=0.8545; lo cual está dentro de los rangos establecidos, que se presentaron en 3.2.

IV. DISCUSIÓN

4.1. Comparar las metodologías clásica, bayesiana y de redes neuronales artificiales.

Al revisar la información se puede apreciar que la metodología clásica, básicamente estadística, no es única para el caso de superficie multirespuestas.

ta. Teniendo como punto inicial, en la época actual, la metodología desarrollada por Derringer y Suich, 1980, se presentan una serie de mejoras que van desde los aportes de Khuri y Cornell, hasta los de Del castillo y otros en la década del 1990 y comienzos del 2000, metodologías diversas que introducen básicamente el cálculo de ecuaciones de regresión para cada variable de respuesta, donde se consideran diversos modelos de primer y segundo grado y que luego se integran de distintas maneras en una sola gran ecuación ó meta modelo. Este meta modelo es resuelto con el uso de diversos paquetes estadísticos y de optimización, con diversos algoritmos, obteniéndose un conjunto óptimo de valores de las variables de entrada x , para valores de salida y y deseados.

Por otro lado la Metodología Bayesiana para optimización multirespuesta, tampoco es única, hay diversos aportes donde destacan los trabajos de Miró et al, 2002, y de del castillo, 2007, así como una vasta lista de autores que dan diversas variaciones produciendo mejoras constantes. El uso de paquetes estadísticos y de optimización es muy importante en estos casos, destacando el uso de MATLAB y el software libre R. Sin embargo, la metodología es bastante similar a la clásica multirespuesta, introduciendo los cálculos respectivos del enfoque de inferencia bayesiana. Se llega a un meta modelo que luego se optimiza, buscando un mínimo ó máximo; que permita obtener la mejor aproximación a un conjunto de valores de respuesta esperados. En estos casos los cálculos tienen bastante minuciosidad y complicación para los cálculos, como se pudo ver en los casos que se presentan de Miró et al y Del castillo. Es indudable que los métodos bayesianos superan a los clásicos, ya que introducen la opinión del investigador, subjetiva, pero además, permiten superar las limitaciones de modelos de parámetros únicos, por modelos que consideran las posibilidades de diversos parámetros de primer y segundo grado para cada una de las variables de entrada y realizando aproximaciones llegan a seleccionar un modelo más adecuado.

En lo que respecta a redes neuronales se tiene los mismos valores de entrada y salida, que para los otros casos, y se aproximan sus relaciones mediante diversos tipos de Redes, destacando los de tipo perceptrón multicapa con algoritmo backpropagation, y los de base radial. La metodología de Tong y Hsieh, 2001, permite obtener resultados similares a los obtenidos por la metodología clásica y bayesiana, con la diferencia que se simplifican los cálculos, a veces complicados para optimizar. Los niveles de error son bastante reducidos. También permiten

obtener los resultados de valores de variables de entrada x que producen un óptimo de variables de salida y .

4.2. Analizar la complementación de las metodologías

Al considerar los tres grandes grupos de metodologías de optimización, considero que el más prometedor es el de redes neuronales, siguiente el método propuesto por Tong y kun-Lin Hsieh, 2001, que consta de los siguientes pasos:

- 1) Preparar los datos de entrada (x) y salida (y) del diseño de experimentos para ser utilizados como datos de entrenamiento y de prueba. Asimismo, tener presente que se requieren tener valores de salida (y_d) deseados, que son distintos de los experimentales.
- 2) Determinar los parámetros óptimos (x), construyendo la RNA 1 utilizando las salidas como entradas y viceversa, para ello se preparan los conjuntos de entrenamiento y prueba respectivos, se entrena la red. Luego en la red entrenada se ingresan los valores que se desea obtener (y_d), como entradas; obteniéndose así valores de las (x_o) variables de entrada optimizadas que permiten obtener las y deseadas.
- 3) Estimar los valores de respuesta (y) con las (x_o) obtenidas en 2). Para ello se construye la RNA 2 con las entradas (x) iniciales y salidas iniciales (y); luego del entrenamiento respectivo se procede a ingresar los valores de (x_o) obtenidos en 2) y así la RNA 2 permite obtener los valores de (y) (de predicción) que deberán compararse con los valores de (y_d) deseados.

Si bien es cierto que la metodología con RNA es más rápida que las metodologías clásica y bayesiana (siempre y cuando se domine las técnicas de RNA y el uso de los software respectivos); cada una de ellas tiene aportes importantes. La primera aproximación de la metodología clásica puede decirse que es fundamental para empezar a conocer el problema, para conocer el tipo de relaciones entre las variables. Seguidamente, la metodología bayesiana, permite introducir los criterios del investigador con mayor minuciosidad y la incertidumbre con relación a los parámetros, así como la obtención de un metamodelo y los valores de las variables de entrada que permiten la optimización. Finalmente con las RNA se obtiene un error de predicción menor que en los otros casos, mayor rapidez en la solución, valores las variables de entrada para la optimización, no se tiene una idea comprensible de las relaciones entre las variables y parámetros.

4.3. Sistematizar un nuevo modelo

De todo lo tratado a lo largo del presente trabajo se llega a la conclusión, que el modelo a seguir para la optimización, luego del diseño del experimento, sería de tres etapas; la primera aplicación de la metodología clásica, la segunda de aplicación de la metodología bayesiana y finalmente la aplicación de la metodología de redes neuronales artificiales. Las tres metodologías se complementan y no considero que se deban eliminar. Sería ideal tener un solo software que pudiera hacer todos los cálculos e integrarlos, para facilitar el trabajo de los profesionales dedicados a trabajos de producción y /o gestión.

V. CONCLUSIONES

De lo realizado en el trabajo se han obtenido las siguientes conclusiones:

1. La optimización debe empezarse con una primera metodología exploratoria, utilizando superficie multirespuesta clásica, buscando un modelo ajustado de primer ó segundo orden; el mismo que después se resuelve encontrando óptimos para las respuestas deseadas.
2. Con la información de la superficie multirespuesta clásica y mayor conocimiento del tema se debe proceder a aplicar la superficie multirespuesta con estadística bayesiana. Obteniendo un conjunto de ecuaciones de respuesta y de dicho conjunto se obtiene los óptimos para las respuestas deseadas.
3. Con la información obtenida en 1) y 2) se procede a construir las redes neuronales artificiales RNA 1 y 2, según se indico en 3.3 y 4.2.; obteniéndose óptimos para los diversos valores deseados que se tengan
4. Las tres metodologías se complementan y contribuyen a conocer mejor los procesos o productos que se quieren optimizar.

VI. REFERENCIAS BIBLIOGRÁFICAS

- [1] Ames, A.E., et al. (1997). "Quality Loss Functions for Optimization Across Multiple Response Surfaces". *Journal of Quality Technology*. vol 29, N° 3
- [2] Del Castillo, E. "Process Optimization. A Statistical Approach". (2007). Editorial Springer. EEUU.
- [3] Derringer, G. y Suich, R. (1980). "Simultaneous Optimization of Several Response Variables". *Journal of Quality Technology*. volo 12, N° 4.
- [4] Khuri, A.I. y Cornell, J.A. (1996). *Reponse Surfaces*, 2da Edición. Editorial Dekker. EEUU.
- [5] Lee-Ing Tong y kun-Lin Hsieh, (2000). "A Novel mean of Applying Neural Networks to Optimize the Multireponse Problem", *Quality Engineering*, Vol.13, N° 1;
- [6] Miro et al (2002). "A Bayesian Approach for Multiple response Surface Optimization in the Presence of Noise Variables" *Journal of Quality Technology*. vol. 34. N° 3.
- [7] Phillips, P. y Kim, K. (1999). "Taguchi Parameter Design with Multiple Quality Characteristics". *Quality Management Journal*. vol6, N° 4.
- [8] Peterson, J. (2004). "A posterior predictive approach to multiple response surface optimization". *Journal of Quality Technology*. vol 36, N° 2.
- [9] Raftery, A, et al (2003) "Bayesian Model Averaging for Linear Regression Models" . *Journal of the American Statistical Association*. vol 92, N° 437.
- [10] Ko, Y., Kim, K. y Jun, C.H. (2005). "A New loss function - based method for multiresponse optimization". *Journal of Quality Technology*. vol 37, N° 1
- [11] Rajagopal, R y Del Castillo, E. (2005). "Model-robust process optimization using Bayesian model averaging" *Techonometrics*. N° 47-2 , 2005
- [12] Pignatiello, J.J. (1993). "Strategies for Robust Multiresponse quality engineering". *IIE Transactions*, vol25, N°3.
- [13] Szu Hui NG. (2010) "A Bayesian Model- Averaging Approach for Multiple- Response Optimization". *Journal of Quality Technology*, vol 42, N°1.
- [14] Vinning, G.G. (1998) "A Compromise Approach to Multiresponse Optimization". *Journal of Quality Tecnology*. vol 30, N°4