

# Uso del análisis envolvente de datos y técnicas de agrupamiento en optimización multicriterio

MATILDE SÁNCHEZ PEÑA\*, MARÍA G. VILLARREAL\*, MAURICIO CABRERA RÍOS\*



La optimización en el terreno de la manufactura se ha caracterizado por el uso de técnicas con un solo objetivo, representa una medida de desempeño (MD). Sin embargo, realmente la mayoría de los problemas en esta disciplina involucran múltiples MDs. Un mayor inconveniente se debe a que estas MDs se encuentran frecuentemente en conflicto. Un ejemplo sobre la contradicción de objetivos para dos MDs en un proceso de manufactura es el de la calidad del producto y su tiempo de ciclo. Frecuentemente, la optimización de la primera MD conlleva a un deterioro en la segunda.

El problema matemático que recoge los elementos de un problema con múltiples MDs en conflicto es el problema de optimización de criterios múltiples. Este problema es denominado en la bibliografía correspondiente *programación de objetivos múltiples*, nombre que surge de las funciones-objetivo que representan a las MDs.

□ El presente artículo está basado en la investigación "Uso del análisis envolvente de datos y técnicas de agrupamiento en optimización multicriterio", galardonada con el Premio de Investigación UANL 2008 en la categoría de Ciencias Exactas, otorgado en sesión solemne del Consejo Universitario, en septiembre de 2008.

La resolución de estos problemas, al contrario de los que comprenden solamente una MD, no busca una solución única, sino un conjunto de soluciones llamadas *soluciones eficientes* o *pareto-eficientes*. Estas soluciones definen una *frontera eficiente*, sobre la cual se encuentran los mejores compromisos alcanzables para ambas MDs.

En la figura 1 se representan soluciones candidatas en términos de sus valores en las MDs

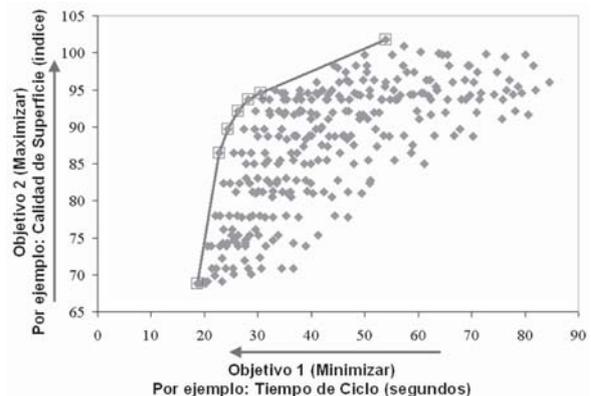


Fig. 1. Representación del problema de optimización de criterios múltiples. Las soluciones candidatas señaladas por rombos se encuentran dominadas por una serie de puntos eficientes -unidos por líneas- que forman una frontera eficiente. Este problema en particular involucra dos criterios u objetivos: minimizar el tiempo de ciclo y maximizar la calidad de la superficie.

\* Programa de Posgrado en Ingeniería de Sistemas, FIME-UANL.

de interés: tiempo de ciclo en segundos que se desea minimizar y calidad, representada por un índice y que se desea maximizar.

Dada la descripción de los objetivos, una solución ideal estaría en la esquina superior izquierda de la gráfica; sin embargo, no existe una solución ahí. Las soluciones eficientes –unidas por líneas para mostrar la frontera eficiente en la figura 1– nunca pueden mejorar en todas las medidas de desempeño al mismo tiempo sin que se pierda en alguno de ellos, mientras que las no eficientes sí. La conveniencia de una representación gráfica se pierde al aumentar el número de MDs en el problema, y por eso se tiene que recurrir a técnicas matemáticas para resolver este tipo de problemas.

La discretización de la región factible que se presenta en la figura 1, propuesta y utilizada por Cabrera-Ríos *et al.*,<sup>18</sup> permite la aplicación de la técnica no paramétrica conocida como Análisis envolvente de datos (AED), la cual se basa en la programación lineal para encontrar la frontera eficiente. La técnica fue desarrollada por Charnes, Cooper y Rhodes.<sup>9,10</sup> Aunque se ha reportado éxito en la utilización del AED, uno de los retos para llevar este método a la práctica tiene que ver con el número de soluciones candidatas que se tienen que evaluar. Este número de soluciones afecta negativamente: 1) la cantidad de veces que el modelo de AED se tiene que resolver, 2) el tamaño del problema que se tiene que resolver. Tener un alto número de soluciones candidatas, sin embargo, es necesario para que la fidelidad de la frontera eficiente que se encuentre sea más alta.

Algunos ejemplos reales del uso de AED para resolución de problemas de criterios múltiples llegan a tener más de 10,000 soluciones candidatas.<sup>2,4,6,8</sup> Esto dificulta la instauración del método en la industria, pues el optimizador MS Excel Solver, que es el más económico del mercado, sólo puede resolver problemas con un máximo de 200 restricciones. La adquisición de software más potente afectaría negativamente el costo de los proyectos de optimización, por lo que lograr una metodología basada en el optimizador mencionado re-

dundería en un beneficio para la industria regional y nacional.

Los resultados de Villarreal y Cabrera Ríos<sup>11</sup> indican que se pueden resolver de manera económica los problemas de criterios múltiples en situaciones reales de la industria a través de la combinación adecuada de métodos de agrupamiento.

En dicho trabajo se comprobó empíricamente que del agrupamiento estadístico de datos para el uso de AED reduce significativamente el tiempo que se tendría que invertir comparativamente con la utilización de los datos sin agrupar. Sin embargo, al reducir el tiempo de ejecución con las medianas de los grupos como medida representativa se sacrificaba también la cantidad de puntos eficientes reales encontrados.

### Planteamiento del problema

En este trabajo se buscan estrategias que permitan mantener la reducción de tiempo de análisis lograda anteriormente, pero aumentando la calidad de la frontera eficiente encontrada. Estas estrategias incluyen la elección de puntos diferentes a la mediana. Por otro lado, se comprueba la eficacia de los métodos realizados en un caso práctico de moldeo por inyección.

### Experimentación y resultados

En esta sección se presentan las metodologías utilizadas de agrupamiento de datos y selección de puntos representativos, así como los resultados asociados a cada estrategia.

En los análisis que se presentan se aprovechó la herramienta construida por Villarreal y Cabrera Ríos,<sup>11</sup> que es una instauración de AED en MS Excel capaz de evaluar 100 vectores a la vez y determinar los vectores eficientes de este conjunto. Se utilizaron 10,000 datos iniciales (vectores en  $R^2$ ), los cuales se obtuvieron aleatoriamente al emplear MS Excel (distribuidos normalmente con media 500 y desviación estándar 20).

Se partió de un agrupamiento por el método de las kmedias, el cual se encuentra disponible en el paquete computacional Minitab. El agrupamiento por k-medias se basa en la mínima distancia euclidiana a los miembros de cada grupo.<sup>12</sup> Tomando los 10,000 datos originales, se obtuvieron 100 grupos (con diferente número de vectores cada uno). La figura 2 ilustra un ejemplo reducido del agrupamiento por este medio, esto es, cinco grupos denotados por diferentes figuras. Con este agrupamiento se seleccionó después un punto por cada grupo como dato representativo, tal como se describe a continuación:

### 1. Medianas de cada grupo como medida descriptiva

Este dato fue utilizado por Villarreal y CabreraRíos; sin embargo, al reducir el tiempo, también se reducía la cantidad de puntos eficientes encontrados exitosamente. Aquí se vuelve a presentar este método para compararlo contra los desarrollos del trabajo presente. Los pasos son:

- 1) Se calculan las medianas de cada grupo en sus dos MDs, y se toma la intersección de éstas como punto representativo de cada uno de los 100 grupos. En la figura 2 se ilustran estos datos representativos para nuestro ejemplo, se señalan con cuadrados vacíos.
- 2) Se analizan los datos con la herramienta de AED y se obtienen los grupos eficientes.
- 3) Los grupos eficientes se dividen en sus componentes.
- 4) El total de vectores componentes son analizados con la herramienta de AED, para llegar a la frontera eficiente. Cuando los componentes son más de 100, se volverán a crear 100 grupos, utilizando k-medias y se iterará desde el paso 1.

### 2. Medias de cada grupo como medida descriptiva

El procedimiento es similar al anterior, sólo que en lugar de las medianas se manejan las medias (representadas con círculos vacíos en la figura 2).

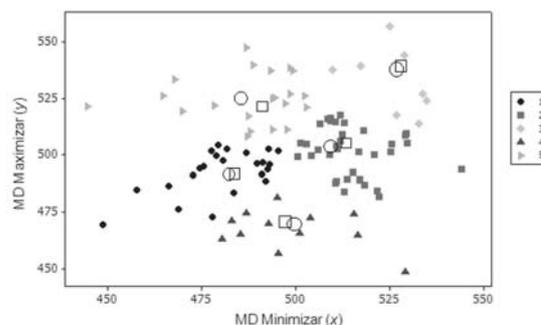


Fig. 2. Ilustración del cálculo de las medianas (cuadrados vacíos) y medias (círculos vacíos) por cada grupo.

### 3. El punto con el máximo valor en la MD a maximizar

En esta variante se toma el punto con el máximo valor en la MD a maximizar como punto representativo de cada grupo, y se sigue el mismo procedimiento que en el caso base involucrando las medianas. En la figura 3, estos puntos representativos se señalan con un círculo vacío y la MD a maximizar se representa con el eje  $y$ .

### 4. El punto con el mínimo valor en la MD a minimizar

Este caso es similar al anterior, pero se toma el valor mínimo en la MD a minimizar como punto representativo. En la figura 3, los puntos de este caso se representan con cuadrados vacíos y la MD se asocia al eje  $x$ .

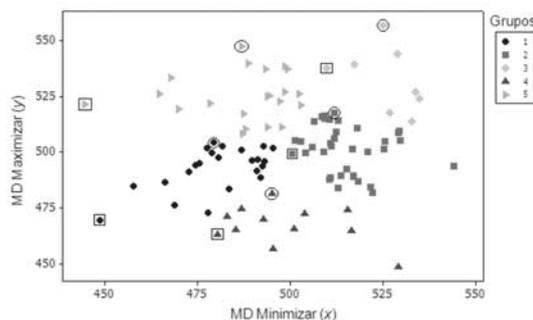


Fig. 3. Selección de los valores mayores en  $y$  (círculos) y valores menores en  $x$  (cuadrados), como representativos en el agrupamiento de datos.

### 5. Combinación lineal convexa del punto con el máximo valor en la MD a maximizar y el punto con el mínimo valor en la MD a minimizar

Con los puntos encontrados en los dos casos anteriores es posible derivar un punto que sea la combinación lineal convexa de ambos, y utilizarlo como punto representativo por grupo. La figura 4 ilustra más claramente los puntos creados; éstos se encuentran en el punto medio de la recta que genera la unión de los datos manejados en los dos métodos anteriores.

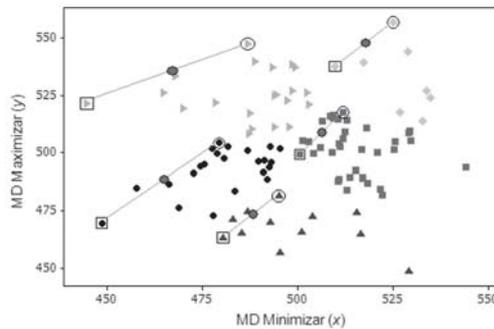


Fig. 4. Representación gráfica del cálculo de la combinación convexa de los datos máximos en y, y mínimos en x.

Tabla I. Resultados comparativos. Se muestran los puntos eficientes encontrados, el porcentaje que éstos representan del total y la cantidad de ejecuciones del AED para cada metodología propuesta.

Metodología	Cantidad de puntos eficientes encontrados en proporción con los puntos de eficientes originales	Cantidad de evaluaciones AED
Medianas de los grupos	3/4	208
Medias de los grupos	3/4	208
Puntos de cada grupo con el máximo valor en y	4/4	171
Puntos de cada grupo con el mínimo valor en x	4/4	200
Combinación convexa de puntos con máximo valor en y y puntos con mínimo valor en x	4/4	220

Los resultados de los cinco casos con estrategias de agrupamiento se presentan en la tabla I. En ésta se ilustra la cantidad de puntos eficientes obtenidos por cada método en proporción con la cantidad de puntos totales pertenecientes a la frontera eficiente, y la cantidad de ejecuciones del AED para llegar a ésta.

Los resultados indican que el uso de puntos representativos no eficientes como la intersección de medias o medianas causa una pérdida de soluciones eficientes, como ocurrió en el estudio anterior. Entonces, elegir puntos eficientes, como el punto con el máximo valor en la MD a maximizar o el mínimo valor en la MD a minimizar, arroja mejores resultados.

### Caso práctico

La aplicación de los métodos aquí presentados se demuestra a continuación por medio de un caso práctico en el terreno de la manufactura. Los datos iniciales utilizados se obtuvieron de una simulación de moldeo por inyección de plásticos<sup>13</sup> llevada a cabo con el programa MoldFlow. Para coleccionar los datos se utilizó un diseño de experimentos de tipo factorial con dos factores: a) temperatura del molde, b) presión de inyección cada uno a nueve niveles. Las medidas de desempeño (MDs) escogidas para este estudio fueron: (1) la presión máxima dentro del molde, (2) el tiempo de ciclo. Se ha observado que estas MDs se entran en conflicto al tratar de minimizar ambas simultáneamente. Es deseable minimizar ambas MDs: 1) para no restar la capacidad de cierre del moldeo, 2) para aumentar la producción. Para efectos del AED, fue necesario transformar linealmente la máxima presión dentro del molde para que correspondiera a un caso de maximización. En el resto de este trabajo, a esta MD transformada se le identifica como Pmax\*.

La parte considerada en este estudio fue la carátula de un celular, la cual se muestra en la figura 5.



Fig. 5. Imagen de la carátula de celular cuyo proceso de inyección se simuló por medio del MoldFlow.

A partir del diseño experimental inicial, se ajustaron metamodelos a cada una de las MDs. La presión máxima dentro del molde se formó por medio de una regresión lineal, y el tiempo de ciclo por una red neuronal artificial con siete neuronas en la capa oculta. Utilizando después estos metamodelos se generaron 10,000 predicciones para el caso práctico.

En primera instancia, los 10,000 datos de las MDs se analizaron exhaustivamente por medio de la herramienta de AED en Excel. Con esto se encontró la frontera real del conjunto, compuesta por cinco puntos, para su comparación posterior. Los puntos eficientes se pueden apreciar en la figura 6. Los valores de los puntos eficientes se especifican en la tabla II.

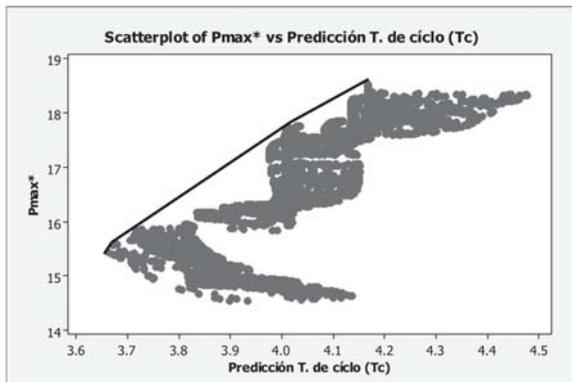


Fig. 6. Cálculo exhaustivo de la frontera eficiente, comprendida por cinco puntos.

Tabla II. Valores de los puntos eficientes

Punto	Tiempo de ciclo	Presión máxima
A	3.6613	15.4427
B	3.6705	15.5752
C	4.0137	17.7225
D	4.0139	17.7232
E	4.1668	18.5385

Los datos fueron agrupados posteriormente por el método de k-medias, como se detalló en la sección de experimentación y resultados. El agrupamiento para este caso se ilustra en la figura 7.

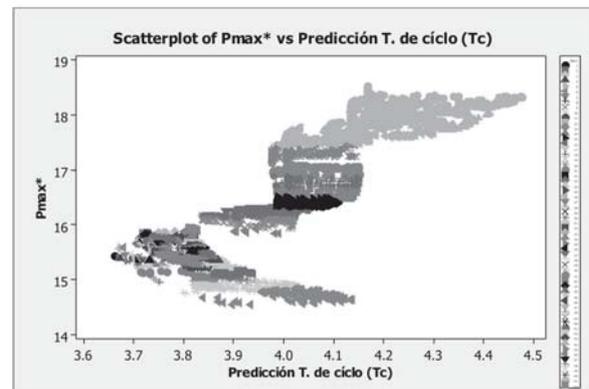


Fig. 7. Agrupamiento original de los datos.

Los puntos representativos que demostraron mayor eficacia en el estudio anterior, esto es la selección de máximos de  $y$ , y la selección de mínimos de  $x$ , se utilizaron en este caso como sigue:

- 1) Se seleccionaron los valores máximos en  $P_{max}^*$  de cada grupo (se recuerda que estos corresponden a los valores mínimos de  $P_{max}$ ).
- 2) Se seleccionaron los valores mínimos de tiempo de ciclo (T. de ciclo) de cada grupo.
- 3) Se analizaron los 200 puntos seleccionados, y se obtuvieron así los grupos eficientes por cada método.
- 4) Se tomaron como eficientes los grupos comprendidos en la unión de conjuntos eficientes por cada punto representativo. Por ejemplo, en la primera iteración, por máximos de  $y$  se obtuvieron los grupos 1, 3, 61 y 96; y por mínimos de  $x$ , los grupos 1, 3 y 61. Así que se

tomaron como eficientes los grupos 1, 3, 61 y 96.

- 5) Se dividieron los grupos en sus componentes.
- 6) Si la cantidad de componentes era menor a 100 elementos, éstos eran analizados en la herramienta de AED una vez más para hacer el cálculo de la frontera real. En cualquier otro caso, los elementos se agrupaban nuevamente por el método de k-medias y se repetían los pasos desde el principio.

Cada etapa del método realizado se reporta en la tabla III y en la figura 8.



Fig. 8. Muestra la evolución de los puntos evaluados en cada tratamiento

Tabla III. Descripción de las diferentes etapas del proceso.

	Cantidad de grupos eficientes.	Cantidades de puntos contenidos en los grupos eficientes	Total de puntos comprendidos para el siguiente agrupamiento
<b>1era Ronda</b>	4	(1, 1, 1298, 1558)	2858
<b>2da Ronda</b>	5	(1, 1, 930, 628, 22)	1582
<b>3era Ronda</b>	5	(1, 1, 80, 115, 22)	219
<b>4ta Ronda</b>	5	(1, 1, 1, 1, 22)	26

El total de ejecuciones que se hicieron del AED con la herramienta de Excel se reporta en la tabla IV.

Tabla IV. Cantidad de ejecuciones necesarias en el método implementado.

Etapa	Ejecuciones
1er Agrupamiento	200
2do Agrupamiento	200
3er Agrupamiento	200
4to Agrupamiento	200
5to Tratamiento (n<100)	100
<b>Total</b>	<b>900</b>

El resultado de este método fue la obtención de la totalidad de los puntos eficientes (cinco puntos), como muestra la figura 9.

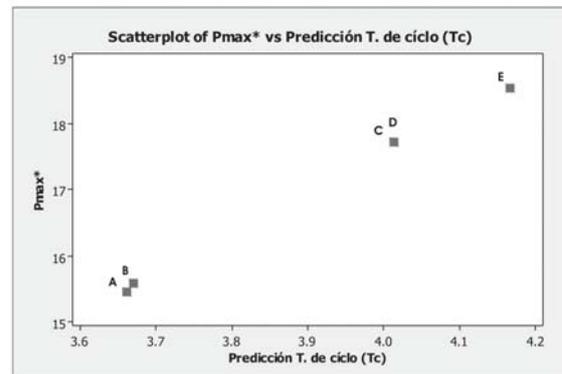


Fig. 9. Los cinco puntos de la frontera eficientes calculada por la metodología planteada.

### Conclusiones y trabajo a futuro

En este trabajo se exploraron métodos que utilizan diferentes puntos representativos, a partir del agrupamiento estadístico de datos, para aumentar la fidelidad de la frontera eficiente en la resolución de problemas de optimización de múltiples criterios con Análisis envolvente de datos. El objetivo principal del estudio se cumplió al mejorar el porcentaje de puntos eficientes reales obtenidos en la primera parte de este trabajo. Se demostró también la aplicación de los métodos estudiados con un caso de estudio de moldeo por inyección.

Dados los resultados, se explorará una integración computacional económica para la adopción de las técnicas presentadas en la industria.

## Resumen

En una primera parte de este trabajo, publicada en *CiENCiAUANL*, se presentaron los resultados preliminares al utilizar diferentes metodologías de agrupamiento de datos para hacer más eficiente el uso del Análisis envolvente de datos para solucionar problemas de optimización de múltiples criterios. Se demostró empíricamente que es factible reducir de forma significativa el tiempo de análisis, si se aplican los métodos propuestos para determinar una frontera eficiente; sin embargo también se reportó un deterioro en la fidelidad de la frontera encontrada. La razón para tal deterioro es que al descartar datos agrupados, se descartaron algunos puntos que sí pertenecían a la frontera eficiente. En este trabajo se exploran la elección mejorada de un punto representativo para cada agrupamiento para aumentar la fidelidad de la frontera encontrada y la adopción de criterios de discriminación para reducir la dimensión del problema multicriterio original.

**Palabras clave:** Optimización multicriterio, Análisis envolvente de datos, Técnicas de agrupamiento de datos.

## Abstract

The first part of this work explored using different Data Clustering Techniques to make the use of Data Envelopment Analysis more efficient to solve Multi-criteria Optimization problems. It was empirically demonstrated that it is feasible to significantly reduce the analysis time when applying the proposed grouping schemes to determine the efficient frontier. Nevertheless, it was also reported that the improvement in the convergence speed implied some deterioration in the fidelity of the efficient frontier. The reason for such deteriora-

tion was that true efficient solutions were discarded when their group was represented by a dominant solution such as the median. In this work the selection of a nondominant representative point for each group is explored to increase the fidelity of the efficient frontier found, and the adoption of discrimination criteria is used to reduce the dimensions of the original multicriteria problem.

**Keywords:** Multiple criteria optimization, Data envelopment analysis, Data clustering techniques.

## Agradecimientos

La realización de este proyecto fue posible gracias al apoyo de la UANL-Paicyt CA-1069-05. Se reconoce también el valioso apoyo Conacyt, la FIME y la UANL en términos de becas a los estudiantes involucrados en el desarrollo de este proyecto.

## Referencias

1. M. Cabrera-Ríos y J.M. Castro, The Balance Between Durability, Reliability, and Affordability in Structural Composites Manufacturing: Preliminary Results, Technical Paper 2003-01-0459, Reliability and Robust Design in Automotive Engineering, 2003 SAE World Congress Book SP-1736, March 2003.
2. M. Cabrera-Ríos, Thesis (PhD), [www.ohiolink.edu/etd/](http://www.ohiolink.edu/etd/), The Ohio State University (2002).
3. M. Cabrera-Ríos, M., K.S. Zuyev, X. Chen, J.M. Castro, and E.J. Straus, Optimizing Injection Gate Location and Cycle Time for the In-Mold Coating (IMC) Process, *Polymer Composites* 23:5 (2002).
4. C. E. Castro, M. Cabrera-Ríos, B. Lilly, J.M. Castro, and C.A. Mount-Campbell, Identifying The Best Compromises Between Multiple Performance Measures In Injection Molding (IM) Using Data Envelopment Analysis (DEA), *Journal of Integrated Design*

- & Process Science, 7, 77-87 (2003).
5. M. Cabrera-Ríos, J. M. Castro, and C.A. Mount-Campbell, Multiple Quality Criteria Optimization In In-Mold Coating (IMC) With A Data Envelopment Analysis Approach, *Journal of Polymer Engineering*, 22:5 (2002).
  6. M. Cabrera-Ríos, J.M. Castro y C.A. Mount-Campbell, Multiple Quality Criteria Optimization In Reactive In-Mold Coating With A Data Envelopment Analysis Approach II: A Case With More Than Three Performance Measures, *Journal of Polymer Engineering*, 24:4 (2004) 435.
  7. M. Castro, M. Cabrera-Ríos y C.A. Mount-Campbell, Modelling and Simulation in Reactive Polymer Processing, *Modelling and Simulation in Materials Science and Engineering*, 12 (2004) S121-S149.
  8. C.E. Castro, M. Cabrera-Ríos, B. Lilly, and J.M. Castro, Simultaneous Optimization of Mold Design and Processing Conditions in Injection Molding, Aceptado para publicación en *Journal of Polymer Engineering* (2005).
  9. A. Charnes, W.W. Cooper, and E. Rhodes, *European Journal of Operational Research*, 2:6 (1978).
  10. A. Charnes, W.W. Cooper, A.Y. Lewin, and L. M. Seiford, *Data Envelopment Analysis: Theory, Methodology, and Applications*. Boston. Kluwer Academic Publishers (1993).
  11. M.G. Villarreal y M. Cabrera-Ríos, Agrupamiento de datos para la solución del problema de optimización multicriterio. *CiENCiA UANL*, 10:2, 137-142 (2007).
  12. J. Han and M. Lamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann Publishers, 335-391 (2001).
  13. M. A. Urbano-Vázquez y M. Cabrera-Ríos, Comparación de diseños experimentales para la predicción de líneas de costura en el moldeo por inyección: Resultados preliminares. *Memorias del Congreso CIINDET* (2006).
  14. M. Ehrgott, *Multicriteria Optimization*, 2nd Edition, Ed. Springer (2005).

### Publicaciones y ponencias derivadas de este trabajo

Los resultados aquí obtenidos se publicaron en la revista internacional *Intelligent Data Analysis*, en las memorias del Quinto Congreso de Innovación y Desarrollo Tecnológico CIINDET 2007 y en la revista *Ingenierías*.

Recibido: 1 de septiembre de 2008

Aceptado: 12 de septiembre de 2008