

Experiment Design for a Multi-objective Partitioning Problem

M. B. Bernábe, J. Vanoye, M. Bustillo, R. González, A. Ochoa and G. Martínez

Abstract— This work presents a factorial statistic experiment for a bi-objective combinatory optimization problem, which optimizes two functions in conflict: geometric compactness and homogeneity for variables from a population problem, which belongs to the Territorial Design (DT) area. This kind of problems invest their biggest effort on the bi-objective clustering to build groups of zones under partitioning properties where the territorial partitions must be as compact and as homogeneous as possible. At this point, the resolution of the compromise between two objectives has to be approached with a multi-objective technique to find non-dominated solutions that at the same time form the set of solutions encompassed in a Pareto Frontier (FP). The method proposed to find the set of non-dominated solutions is supported on basic aspects from the order theory, in particular from the Hasse diagram to obtain the Minima, and to manage the computational cost we have incorporated the Variable Neighborhood Search metaheuristic (VNS). To calibrate the VNS parameters we have employed a factorial experiment known as Box Benhken (BB) and Response Surface, in this way we have achieved an ideal combination of parameters to find satisfactory solutions to the multi-objective problem under study.

Keywords— Compactness, Homogeneity, Experiment Design, Pareto Frontier, Partitioning.

I. INTRODUCCIÓN

LOS PROBLEMAS denominados multiobjetivo, son aquellos que tratan con la existencia de múltiples criterios que se encuentran en conflicto entre ellos. Esto implica que existen diferentes soluciones y es necesario tomar una decisión en base a una serie de criterios contrapuestos. El proceso de tomar una decisión reside en que planteado un problema, debe formarse el conjunto de puntos factibles, o de otro modo, el correspondiente conjunto determina las restricciones del problema para asociarle a cada alternativa, criterio u objetivo, un grado de deseabilidad. Posteriormente, se rastrea mediante alguna técnica, un conjunto de posibles soluciones alternativas. Estas posibles soluciones son las que satisfacen las restricciones y preferencias, las cuales se ejecutan sobre los objetivos planteados. Estos objetivos son funciones, las cuales deben ser optimizadas y representan al problema matemáticamente.

M. B. Bernábe, Facultad de Ciencias de la Computación, BUAP, Puebla, México, beatriz.bernabe@gmail.com

J. Vanoye, Universidad Autónoma de Hidalgo, Hidalgo, México, jorge@ruizvanoye.com

M. Bustillo, Facultad de Ciencias de la Computación, BUAP, Puebla, México, bustillo@cs.buap.mx

R. González, Facultad de Ciencias de la Computación, BUAP, Puebla, México, rogelio.gzzvzz@gmail.com

A. Ochoa, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, México, alberto.ochoa@uacj.mx

G. Martínez, Facultad de Ciencias de la Computación, BUAP, Puebla, México, gguzman@cs.buap.mx

El problema que tratamos en este trabajo resuelve el agrupamiento de dos objetivos para datos geográficos en el plano R^2 , lo que implica obtener no solo un costo como sucede con el particionamiento clásico. Este agrupamiento multiobjetivo debe generar un conjunto de soluciones, donde un costo es compacidad y otro es homogeneidad. En este punto, el uso de particionamiento como un método de optimización con VNS, es muy frecuente en diseño territorial, el cual busca la compacidad y balanceo para alguna variable descriptiva.

Muchos son los esfuerzos para la optimización de un solo objetivo [1, 2], sin embargo, la necesidad de incluir otra medida de calidad es rebasada por la demanda de problemas reales.

En este trabajo, en un contexto multiobjetivo, una propuesta para particionamiento alrededor de los medoides es presentada. Se resuelve un problema que involucra variables de tipo censal y pertenece a la categoría de diseño de territorio, el cual es de alta complejidad computacional [3]. Dos medidas de calidad para particionamiento son optimizadas simultáneamente en el proceso de clustering con la metaheurística VNS. La primera función obedece a un concepto geométrico de distancias y la segunda se centra en un cálculo de balanceo para una variable descriptiva sobre indicadores de la población con problemas económicos, de este modo, en el particionamiento implícito que hemos desarrollado, destaca una ventaja sobre los clásicos agrupamientos k-medias y k-medoides: la adición de otra función de costo que actúa sobre vectores de variables. Los resultados se muestran en el Frente de Pareto que se consigue con la Mínima (soluciones no comparables y no dominadas) [4] y la aproximación se realiza con VNS.

Finalmente se calibran los parámetros de ambos objetivos con la técnica de Diseño de Experimentos factorial Box Behnken y Superficies de Respuestas para encontrar un balance entre ellos [5].

El presente trabajo se encuentra organizado como sigue: Esta introducción como sección 1. En la sección 2 se exponen aspectos básicos de la teoría multiobjetivo y la descripción del problema para continuar con la sección 3, donde se describe la modelación del problema de particionamiento multiobjetivo. En la sección 4 se describe el problema en la praxis de un problema socio-económico. La sección 5 se ocupa de presentar la teoría necesaria sobre diseño de experimentos para que finalmente en la sección 6 se describa el experimento estadístico aplicado al problema multiobjetivo. La última sección corresponde a las conclusiones y trabajo posterior.

II. PRELIMINARES MULTI OBJETIVO

Un problema de programación multiobjetivo responde a la siguiente formulación:

(a) Optimizar $(f_1(x), f_2(x), \dots, f_p(x))$ sujeta a $x \in X$

Donde $x = (x_1, x_2, \dots, x_n)$ son las variables de decisión, X es el conjunto de soluciones f_i son cada uno de los objetivos y $(f_1(x), f_2(x), \dots, f_p(x))$ es la función vectorial con $Y = f(X)$ que es el espacio de objetivos o espacio imagen.

Para realizar las comparaciones entre vectores se utilizaron los siguientes órdenes, donde la relación se establece atendiendo al valor que tomen los objetivos en dichos puntos:

1) Orden de Pareto: Donde un punto x es preferido a x' si se verifica que:

(b) $f_i(x) \geq f_i(x') \forall i = 1, \dots, p$ con al menos un j tal que $f_j(x) > f_j(x')$

2) Orden débil de Pareto: Donde un punto x es preferido a x' si se verifica que:

(c) $f_i(x) \geq f_i(x') \forall i = 1 \dots p$

El orden de Pareto y el orden débil de Pareto son ordenes parciales y nos establecen que una combinación será preferida a otra siempre que en el primer caso mejore a todos los objetivos, mejorando a uno de ellos de forma estricta, y en el segundo, denominado débil, siempre que mejore todos los objetivos estrictamente [6].

Los puntos anteriores indican que no existirá una combinación que sea a la vez solución de todos los objetivos. De esta forma el primer concepto que perdemos es el de óptimo tal y como se entiende en la programación mono objetivo tradicional, y lo que buscaremos para solucionar nuestro problema serán las denominadas soluciones eficientes. La generalización de óptimo da lugar al concepto de eficiencia, el cual se puede definir para maximización, utilizando la ordenación de Pareto de las dos formas siguientes [6]:

Definición 1. Un punto x^* es eficiente si no existe x tal que x sea preferido a x^* . Si se utiliza el orden Pareto x^* es eficiente si no existe x tal que

$f_i(x) \geq f_i(x^*) \forall i = 1, \dots, p$ con al menos un j tal que $f_j(x) > f_j(x^*)$.

Definición 1.1 Un punto x^* no es eficiente si existe x tal que x es preferido a x^* . Es decir, x^* no es eficiente si existe x tal que $f_i(x) \geq f_i(x^*)$ para algún $i = 1, \dots, p$

En general, existe más de una solución eficiente. En este trabajo, la selección entre las diferentes soluciones no comparables que son producto de un compromiso entre la compacidad geométrica y la homogeneidad de las variables descriptivas es dejada al sistema de toma de decisiones final que de acuerdo a los intereses del proyecto se pueden seleccionar una de las soluciones que se enmarcan en el Frente de Pareto. Las soluciones que nuestro algoritmo genera son no comparables y no dominadas. En este punto, una notación útil para entender la dominancia es la siguiente:

Definición 2. Un problema multiobjetivo puede definirse en el caso de minimización (y análogamente para el caso de maximización) como:

Minimizar $f(x)$ dado que $f : F \subseteq R^n \rightarrow R^q$ $q \geq 2$ con región factible en $A = \{a \in F: g_i(a) \leq 0, i = 1, \dots, m\} \neq \emptyset$

El conjunto A se denomina región factible y se dice que el problema se encuentra sujeto a las restricciones $g_i: R^n \rightarrow R$ que son funciones cualesquiera.

En la optimización multiobjetivo se tiene que decidir un cierto esquema de mejoría de una solución sobre otra, frecuentemente conocido como esquema de dominación y su definición se basa principalmente en que la solución de un problema multiobjetivo no es única y por lo tanto el tomador de decisiones debe elegir de entre una gama de posibles soluciones que no se pueden mejorar entre sí, es decir que no se dominan. La claridad de este concepto, se da cuando pensamos que dentro del campo de los números reales se encuentra definido el orden de manera natural. Para R^n , es posible extender el concepto mediante la siguiente definición.

Definición 3. Dados x, y vectores en R^n $x \leq y$ si y solo si $x_k \leq y_k$ para todo $k \in \{1, \dots, n\}$ y $x < y$ si y solo si $x \leq y$ con $x \neq y$, donde \leq es el orden usual en R .

A. Frente de Pareto

La relación de dominación es conocida dominación de Pareto definida como sigue:

Definición 4. Dado el problema multiobjetivo, minimizar $f(x)$, donde $f : F \subseteq R^n \rightarrow R^q$, $q \geq 2$ con $A \subseteq F$ la región factible. Decimos que un vector $x^* \in A$ es no dominado o un óptimo Pareto si no existe un vector $x \in A$ tal que $x < x^*$ [7, 8].

Así, la respuesta al problema de hallar las mejores soluciones (las soluciones no dominadas, como quiera que se defina la dominación dentro de la técnica) en un problema multiobjetivo es a lo que se le llama el conjunto solución del problema y el conjunto de valores de la función objetivo con dominio restringido a los vectores del conjunto solución (es decir, los vectores no dominados) es lo que conocemos como Frente de Pareto. De esta manera, el concepto de conjunto de vectores no dominados conduce lógicamente al concepto de conjunto parcialmente ordenado.

Definición 5. El conjunto $E(A; f)$ de soluciones de Pareto eficientes (también conocido como conjunto de óptimos de Pareto) se define de la manera siguiente:

$E(A, f) = \{a \in A: \nexists b \in A \text{ que cumpla } f(b) < f(a)\}$

Es decir, el conjunto de todos los vectores no dominados bajo el esquema de Pareto.

Un concepto estrechamente relacionado con el Frente de Pareto es el de óptimo de Pareto. Tanto el óptimo de Pareto como el Frente de Pareto son el marco sobre el que se trabaja dentro de la toma de decisiones multicriterio. El conjunto de óptimos Pareto para un problema multiobjetivo dado, es un conjunto parcialmente ordenado (poset) visto formalmente. En los problemas multiobjetivo se buscan los elementos minimales del espacio de solución R^n visto como un poset con la relación \leq dada en la definición 4 [7, 8].

B. Descripción del problema

La meta es obtener una partición de datos espaciales (Unidades Geográficas UG) cuya composición está dada por 2 componentes: coordenadas geográficas en el plano R^2 y un vector de características descriptivas de carácter censal (variables poblacionales). La primera componente permite que se obtenga una matriz de distancias para el proceso de cálculo de la compacidad geométrica, una de las funciones de objetivo a minimizar. El vector de descripción se utiliza para optimizar la segunda función objetivo y consiste en minimizar la homogeneidad de las variables censales. La selección de

variables se ha hecho respetando los datos y sus valores que corresponden a la población económicamente no activa del censo de población y vivienda INEGI [9].

De acuerdo al INEGI, población económicamente no activa o inactiva (PEI) incluye a las personas de 12 o más años que en la semana de referencia no participaron en actividades económicas, ni eran parte de la población desocupada. Esta se clasifica en:

Población no económicamente activa disponible. Son personas de 12 y más años que no trabajaron ni tenían empleo y no buscaron activamente uno, por desaliento o porque piensan que no se los darían por la edad, porque no tienen estudios, etc.; pero estarían dispuestas a aceptar un trabajo si se les ofreciera, sin embargo no buscan activamente uno.

Población no económicamente activa no disponible. Son personas de 12 y más años que en la semana de referencia no trabajó ni tenía empleo, y tampoco buscó activamente uno y no estaría dispuesta a aceptar un trabajo aunque se le ofrecieran, es decir es la población no disponible para ingresar al mercado de trabajo por dedicarse a: los quehaceres del hogar, estudios o ser jubilados o pensionados, estar incapacitados para trabajar o pertenecer al grupo de otros inactivos, como: ociosos voluntarios, alcohólicos, etc.

Suponiendo que existe un programa de atención a esta porción poblacional, se requieren agrupaciones de las zonas de población económicamente inactiva para su análisis y así proponer programas para atender este sector. La partición que se busca es un conjunto de clases donde sus elementos estén muy cercanos geográficamente, y las UG dentro de los grupos sean determinadas por variables de actividad no económica y además balanceadas para la variable censal “población económicamente inactiva”.

Este problema se entiende también como bi-objetivo de particionamiento geográfico y consecuentemente es de alta complejidad computacional [3]. Su naturaleza combinatoria exige para su solución el uso de métodos heurísticos con el fin de lograr soluciones aproximadas y al mismo tiempo la inclusión en la implementación de herramientas multiobjetivo para encontrar un conjunto de pares de soluciones no dominadas y no comparables. La Mínima ha sido un método efectivo que contribuyó teóricamente para obtener dichas soluciones en el Frente de Pareto y por otro lado para comparar la eficiencia del FP que hemos encontrado. Para el manejo del costo computacional se prefirió a VNS descendente [10].

III. MODELACIÓN

Existe un espacio físico de búsqueda para la agrupación geográfica. Las unidades geográficas (UG) son finitas, esto es, cada elemento es representado por su ubicación espacial y por un arreglo de variables descriptivas de carácter censal. El problema es discreto, combinatorio, y entero-binario. La agregación se realiza bajo las propiedades de particionamiento (agrupamiento sobre medoides) [8].

Para la compacidad, se resuelve la agrupación de tal forma que las UG que componen los grupos estén cercanas geográficamente, entonces, la función objetivo que minimiza la suma de las distancias entre los UG resuelve la compacidad. Por el lado de la homogeneidad, se optimiza buscando un equilibrio entre la variable censal de interés.

El proceso consiste en que una vez formados los grupos bajo la minimización de distancias, se calcula la homogeneidad a la agrupación creada dado que en problemas multiobjetivo, la función a optimizar tiene el mismo dominio para todos los objetivos [11]. De este modo sobre una misma partición se optimiza la compacidad y la homogeneidad, entonces se elige la mejor alternativa respecto a los m objetivos. Matemáticamente, existe un conjunto X que es subconjunto del espacio R^n tal que

$$f_i: X \rightarrow R^n, i = 1, \dots, m \text{ donde existen } m \text{ objetivos [11]}$$

A. Formulación matemática

Sea UG un dato espacial definido por sus componentes en el espacio y una descripción de variables censales dada por un vector.

Notación:

UG =unidad geográfica básica,

GT =Grupo Territorial,

m = número de GT , n =número de UG tal que

$m < n$, i =índice de grupo territorial, j =índice de unidad geográfica básica,

C_i =centroide del i -ésimo grupo territorial GT_i .

El modelo en cuestión es entero mixto y hace uso de las variables binarias para modelos de este tipo.

1) Formulación para la selección de las variables censales y sus cotas:

A es el conjunto de atributos cuantificables del cual será seleccionado un subconjunto de acuerdo al problema.

VA_{kj} es el valor del k -ésimo atributo contenido en la j -ésima UG_j .

α_k, β_k parámetros de tolerancia para VA_k en cualquier UG .

$(VA_{kj}) = \sum_{j=1}^n VA_{kj} X_{ij}$ es el valor para el k -ésimo GT .

X_{ij} variable de decisión.

$\overline{VA}_k = i/m \sum_{j=1}^n VA_{kj}$ es el valor meta para el k -ésimo atributo en cualquier UG .

2) Formulación para minimización de distancias:

$d_{ij} = d(C_i, UG_j) X_{ij}$ es la distancia de la j -ésima UG en el i -ésimo GT respecto a su centroide.

3) Formulación para los dos objetivos

El cálculo de distancias entre UG puede expresarse como:

a) $D_{ij} = \sum_{j=1}^n d(C_i, UG) X_{ij}$ y el valor de homogeneidad para variables se expresa como:

b) $H_{variable\ UG} = \sum_{j=1}^m (\overline{VA}_k - VA_{ki})$ Considerando a) y b) implica:

c) Minimizar $y = f(x) = (f_1, f_2)$.

Donde f_1 es el costo de minimizar la distancia entre UG de acuerdo a la ecuación a) que debe formularse como función, y f_2 es el costo de minimizar la homogeneizar de una variable censal de las UG . Esta función se debe expresar a partir de la ecuación b).

Las funciones f_1, f_2 como en c) están sujetas a: $GT_i = \emptyset$ para $i = 1 \dots k$ (los grupos no son vacíos), $GT_i \cap GT_j = \emptyset$ para $i \neq j$ (No existen UG repetidos en distintos grupos), $\bigcup_{i=1}^k GT_i = UG$ (la unión de todos los grupos son todos los

UG), $\alpha_k \leq VA \leq \beta_k$ (cotas para las variables), $\sum_{i=1}^m X_{ij} = 1$ es la asignación de UG $X_{ij} = 1$ si $UG \in GT_i$ o $X_{ij} = 0$ si $UG \notin GT_i$ (variables de decisión), $y = (y_1, y_2) \in Y \subset R^2$ es el vector objetivo.

Dado que a) y b) son las ecuaciones implícitas en las funciones objetivos, el propósito ahora es presentar a ambas como funciones a optimizar en el proceso de partición con VNS. La heurística VNS encuentra una solución inicial aleatoria (solución actual). Esta solución es una partición donde se ha minimizado compacidad y homogeneidad obteniéndose un par (c_i, h_i) . Se genera la siguiente solución (c_{i+1}, h_{i+1}) y se compara con la anterior de acuerdo al orden Pareto. Si esta nueva solución (c_{i+1}, h_{i+1}) es no comparable y no dominada con respecto a (c_i, h_i) , se etiqueta como solución “subóptima” y se convierte a solución actual para compararse con la “próxima”. Este proceso se repite hasta que los parámetros de VNS lo permitan obteniéndose así el conjunto de soluciones no dominadas que forman el Frente de Pareto.

El orden Pareto no comparable es un orden que se apoya del orden Pareto para generar soluciones no comparables y no dominadas, es decir, un orden Pareto implica:

Dada una solución (a, b) la siguiente solución (a', b') es aceptada si y solo si

$$(a' > a \wedge b' = b) \vee (b' > b \wedge a' = a) \vee (a' > a \wedge b' > b) \vee (a = a' \wedge b = b') \quad (1)$$

Cuando la comparación del par de soluciones se realiza mediante una expresión, solo un punto en la frontera de Pareto se alcanza.

La negación de la expresión (1) permite la producción de aproximaciones a la frontera de Pareto por varias cadenas como en el diagrama de Hasse que a su vez es la Mínima [4]. Sin embargo, también es necesario examinar iterativamente que las soluciones no comparables satisfagan con la dominancia de Pareto. Por último, el conjunto solución obtenido es un conjunto de soluciones no dominadas y no comparables (Frontera de Pareto). Un par de soluciones no es comparable en un determinado orden parcial si no cumple con la propiedad de tricotomía, en este caso significa que dos pares (a, b) y (a', b') son Pareto no comparables si $\neg((a, b) < (a', b')) \wedge \neg((a', b') < (a, b))$, esto es $(a > a' \vee b > b') \wedge (a' > a \vee b' > b)$ (2)

Bajo este orden parcial estricto que hemos llamado Pareto no comparable, y combinado convenientemente con el orden Pareto, obtenemos el conjunto de soluciones no dominadas [12].

Para asegurar que las soluciones que obtenemos son un conjunto de soluciones que forman el Frente de Pareto, se ha considerado la aplicación Nodom, la cual encuentra soluciones minimales de un conjunto de pares de soluciones dado, Nodom (2007). Se comparan así las soluciones generadas del método que aquí se expone contra las que Nodom encuentra. Recordemos que los minimales son un conjunto de puntos no comparables y no dominados en un diagrama de Hasse [4, 12].

IV. APLICACIÓN MULTI OBJETIVO

Cuando un proyecto de tipo económico-social de la población es propuesto, son diversos los aspectos a considerar: estructura organizacional, plan del proyecto, evaluación presupuestal etc. Para este tipo de problemas se propone abordarlo con un esquema de agrupamiento de zonas tales que cumplan con restricciones del problema. En este punto, nos centramos en ofrecer un conjunto de grupos de población económicamente no activa que refleja diferentes escenarios. Estos grupos están formados por objetos espaciales o unidades geográficas conocidas como Areas Geoestadísticas Básicas (Agebs). Cada Ageb está constituido por un vector de variables que obedecen a variables censales.

Un factor importante es tomar la decisión sobre la viabilidad de asignación de recursos sobre un sector de la población vulnerable. Se ha elegido a la población económicamente no activa como caso de estudio. Asumiendo que existe un programa gubernamental para impulsar acciones de apoyo a este sector de la población, se requiere un conjunto de agrupaciones que expongan la distribución de esta población en cuanto a las siguientes variables censales, las cuales están fuertemente correlacionadas: 1) Población de 12 años y más económicamente inactiva que es estudiante, 2) Población desocupada, 3) Población de 12 años y más económicamente inactiva que se dedica a los quehaceres del hogar y 4) Población económicamente inactiva [9].

A. Descripción de la aplicación

El modelo que hemos expuesto ha sido aplicado a un problema social-económico, donde se asume que se requiere de 8 agrupaciones compactas donde los Agebs se encuentren muy cercanos entre ellos con el fin de facilitar los traslados y mantener balanceo entre la población de interés, es decir, los grupos también deberán estar integrados por las 4 variables poblacionales económicamente no activas. Se mantiene homogeneidad en poblacionalmente inactiva.

Se muestra como ejemplo el Frente de Pareto de la prueba 4 que se eligió de entre 10 pruebas para 8 grupos. Los valores de VNS son de 15 para búsqueda local y 2 para estructura de vecindad. [1, 10].

Para las 10 pruebas realizadas, el valor de homogeneidad se mantuvo estable y el tiempo de respuesta osciló entre 31 y 35 segundos. En la Fig. 1 se nota claramente una intersección entre las soluciones logradas con nuestro algoritmo y las soluciones filtradas con Nodom [13]. Este resultado indica que el método construido genera todas las soluciones no dominadas y algunas adicionales (un tanto cuestionables), sin embargo, estas soluciones fuera del frente de Pareto están muy cerca del conjunto de soluciones no dominadas, lo que puede favorecer la selección de alternativas en la toma de decisiones multicriterio.

En el gráfico Fig. 1 el eje de las x representa la compacidad contra la y que es la homogeneidad. Si al decisor le interesa sacrificar compacidad u homogeneidad deberá escoger la solución de entre los puntos señalados. Los valores en cada tabla asociada al gráfico son los costos del valor mínimo para ambos objetivos.

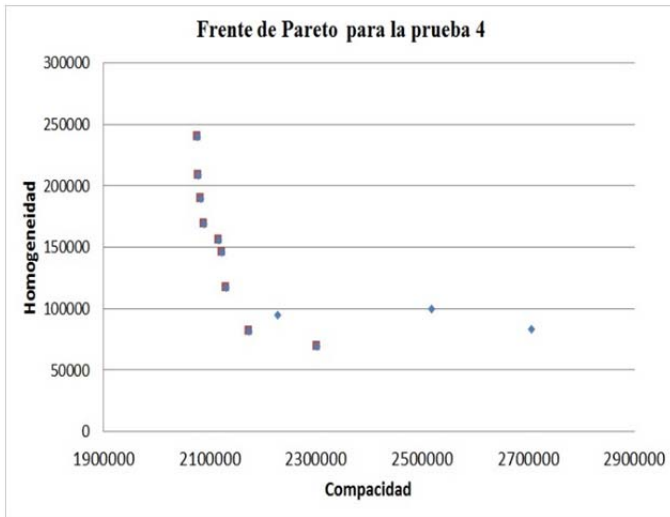


Figura 1. Frente de Pareto para el problema población económicamente inactiva con VNS.

B. La Mínima para el problema multiobjetivo

De las diferentes pruebas realizadas fue posible comprobar que nuestro modelo multiobjetivo y el asociado algoritmo con VNS, genera resultados prometedores. Para fines ilustrativos, hemos elegido una prueba de 8 grupos como se muestra en Fig. 1 con su respectivo Frente de Pareto. Los valores de las coordenadas (soluciones no dominadas y no comparables), es el conjunto Mínima: Minimales) y se disponen en la Tabla I [4].

TABLA I
SOLUCIONES NO DOMINADAS

Soluciones	Minimales para Población Económicamente Inactiva	
	<i>compacidad</i>	<i>homogeneidad</i>
1	2130871	117075
2	2078086	209036
3	2076279	240148
4	2518534	100088
5	2089252	168654
6	2228036	95155
7	2706083	83106
8	2123688	145806
9	2082765	189411
10	2173516	81880
11	2116561	155758
12	2302591	69119

Doce son las soluciones que generó la prueba que estamos presentando. Los valores en rojo son las soluciones que coinciden con Nodom y las soluciones azules son cercanas al frente de Pareto.

Cuando debe tomarse una decisión, la disyuntiva mayor reside en cual o cuales de las 12 soluciones es la mejor. La tarea es difícil y existen algunas técnicas multicriterio para ello, sin embargo, de acuerdo al problema que hemos resuelto, es posible elegir aquella con alta compactidad y poca homogeneidad si los recursos de traslado son acotados. Con el propósito de ser imparcial en los dos criterios, la prueba que

hemos presentado la elegimos en el punto de inflexión entre compactidad y homogeneidad.

Con el fin de comparar la calidad de las soluciones y sobre todo la eficiencia de nuestro algoritmo, hemos recurrido a una aplicación que selecciona de entre un conjunto de soluciones, aquellas que son no dominadas Ndominated [13]. En este punto, la contribución de este trabajo es considerable ya que nuestro algoritmo de partición multiobjetivo con VNS identifica soluciones no dominadas al mismo tiempo que va construyendo los grupos compactos y homogéneos. Ndominated solo acepta un vector de tamaño limitado y su función es filtrar datos no dominadas, no realiza ni un procedimiento adicional. Por otra parte, encontramos que con algunos paquetes estadísticos se consigue agrupar datos multivariados pero con la notable desventaja que solo un criterio es optimizado ofreciendo óptimos locales. La comparación en este escenario no es viable dado que nuestro algoritmo optimiza 2 objetivos. Finalmente un programa competitivo que considera 2 funciones a minimizar con Búsqueda Tabú (BT), fue usado con los mismos datos Agebs y un FP fue obtenido. Sin embargo la medida de calidad para homogeneidad es diferente para este caso [19]. Comparar el FP con BT contra FP con VNS es un desafío, pero en este trabajo no se dispone de suficiente información sobre las soluciones no dominadas con BT. Esto implica que la comparación sobre la calidad de las soluciones de ambos conjuntos no es consistente. Por otra parte el programa multiobjetivo con BT no proporciona información sobre el método que construye el FP, de tal forma que hasta el momento, multiobjetivo con VNS indica que es mejor en cuanto a los resultados que genera, no solo por la complejidad en el programa donde están enlazados los mecanismos de agrupamiento con VNS y la identificación de la Mínima de forma simultánea, también por que las soluciones bien enmarcadas en el FP con sus respectivos valores. Por último, un mapa es creado con nuestro programa y ahí pueden analizarse los grupos de la partición. Pero más lejos de esta comparación, es necesario ocupar técnicas exactas para comparar FP, lo cual es un trabajo complicado y merece ser tratado en un artículo posterior.

Con el fin de observar el FP con BT, a continuación presentamos el gráfico para 8 grupos con población económicamente inactiva. En Fig. 2 se aprecia el FP sin otra información adicional debido a que el programa disponible requiere otros procedimientos que reflejen los datos de la Mínima y sobre todo el método que la obtiene. El FP con BT se distingue pero no pudimos tener acceso a la información relacionada con las soluciones dominadas [19]. Con este resultado y el que nosotros hemos alcanzado con VNS, se asume que una discusión prometedora se anticipa cuando se confronten los dos FPs (VNS, BT).

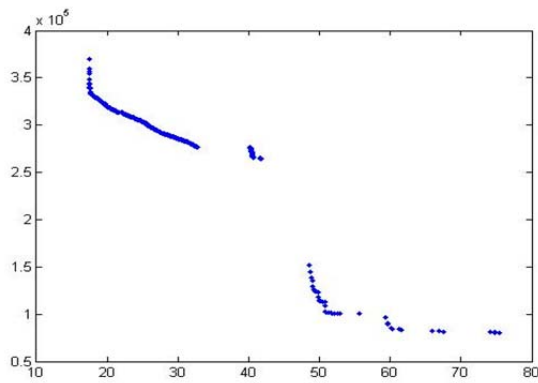


Figura 2. Frente de Pareto para el problema población económicamente inactiva con BT.

La siguiente Fig. 3 describe la solución gráfica asociada con la Fig. 1 y la Tabla I de nuestra aplicación multiobjetivo con VNS y la Mínima. El mapa es el resultado de integrar el código de programación del problema multiobjetivo con un sistema de información geográfico SIG [14]. El resultado es alentador: la compacidad es aceptable a pesar de la inclusión de la homogeneidad de las variables socio-económicas sobre población económicamente inactiva.

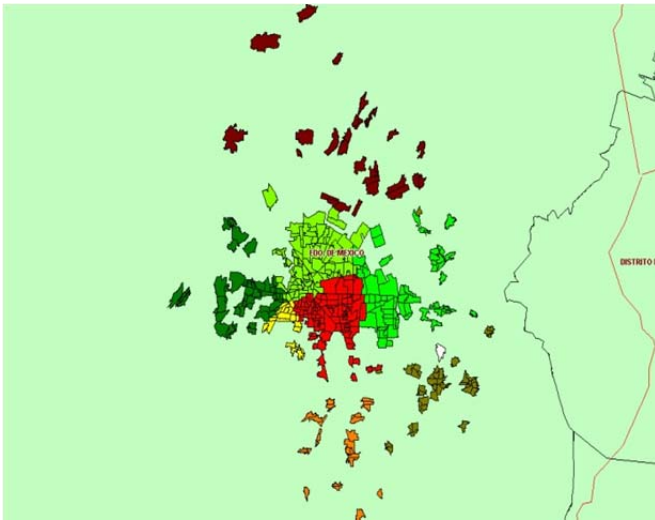


Figura 3. Gráfico de 8 grupos sobre compacidad-homogeneidad para el problema población económicamente inactiva con VNS.

V. DISEÑO DE EXPERIMENTOS

Los parámetros sometidos a las pruebas de la sección anterior obedecen a trabajos previos donde se han calibrado los valores de búsqueda local y estructura de vecindad de VNS para el costo de compacidad [1], pero para nuestro problema con dos objetivos, esta calibración no pudo ajustarse debido a la presencia adicional del costo de homogeneidad. La implicación más simple para proponer una calibración que pueda dar lugar a formular un diseño de experimentos para 2 objetivos, consistió en dar valores aleatorios a partir de los ya conocidos en otros trabajos [1]. Más adelante somos explícitos en la combinación de los parámetros de acuerdo al experimento factorial Box Bhenken [5, 15]. De este modo, hemos aplicado la técnica de diseños de experimentos factoriales Box Bhenken y Superficies de Respuesta para la

encontrar una calibración de parámetros de VNS: Búsqueda Local (LS, Local Search por sus siglas en inglés) y Estructuras de Vecindad (NS Neighbourhood Structures, por sus siglas en inglés) [10]. Esta notación es la usual en las siguientes secciones.

A. Preliminares de Diseño de Experimentos

El diseño experimental estudia procesos, los cuales pueden verse como una caja negra donde ingresan diversas variables que interactúan para producir un resultado. Las variables que ingresan al proceso se denominan variables de entrada (factores) y el resultado, variable de salida (respuesta). El nivel de la variable de salida depende de los niveles que adopten las variables de entrada. La búsqueda de combinaciones óptimas de las variables de entrada da lugar al diseño experimental, el cual es una prueba (o un conjunto de pruebas) durante la cual se realizan cambios sistemáticos y controlados a las variables de entrada para medir el efecto sobre la variable de salida óptima [5]. El diseño experimental utiliza técnicas como la regresión múltiple, la superficie de respuesta, el diseño de parámetros y varias extensiones del análisis de varianza, así como los análisis gráficos y las comparaciones múltiples.

Es frecuente que en muchos procesos existan varios factores de los que es necesario investigar de manera simultánea su influencia sobre una o varias variables de respuesta, donde cada factor tiene la misma importancia a priori desde el momento que se decide estudiarlo, y es poco justificable suponer de antemano que los factores no interactúan entre sí.

El objetivo de un diseño factorial es estudiar el efecto de varios factores sobre una o varias respuestas o características de calidad y determinar una combinación de niveles de los factores en la cual el desempeño del proceso sea mejor que en las condiciones de operación actuales; es decir, encontrar nuevas condiciones de operación del proceso que eliminen o disminuyan ciertos problema de calidad en la variable de salida.

Un diseño de experimentos factorial o arreglo factorial es el conjunto de puntos experimentales o tratamientos que pueden formarse considerando todas las posibles combinaciones de los niveles de los factores.

Por otra parte, La Metodología de Superficie de Respuesta (MSR), es una colección de técnicas que permite al investigador inspeccionar una respuesta, que se puede mostrar como una superficie, cuando los experimentos investigan el efecto que tiene el variar factores cuantitativos en los valores que toma una variable dependiente o respuesta [15]. Esto es, se trata de encontrar los valores óptimos para las variables independientes que maximizan, minimizan o cumplen ciertas restricciones en la variable respuesta.

La tendencia en el desarrollo de la metodología para superficies de respuesta ha sido la construcción de diseños experimentales compactos, con un mínimo de unidades experimentales: el investigador se concentra en las propiedades de los estimadores de todos los parámetros de la función de respuesta, propiedades que dependen del diseño empleado.

Para estimar una superficie de respuesta, los modelos lineales de orden menor o igual a tres se han empleado con

frecuencia por su sencillez y fácil interpretación; sin embargo Cady y Laird (1973) mencionan que los polinomios de grado fraccionario (modelos de los pseudocuadráticos que mediante una transformación lineal se traducen a cuadráticos) [16]. Estos pueden dar una mejor aproximación en algunos experimentos. Por ejemplo, en este artículo, el modelo cuadrático proporcionó un mejor ajuste.

Las bases de la metodología para superficies de respuesta se derivan de la teoría del modelo lineal general. Se asume que existe una variable de respuesta que depende de variables independientes cuantitativas a través de una función f que puede ser complicada o desconocida. La función usualmente se aproxima en la región de interés por un polinomio de orden bajo generalmente menor o igual a tres o pseudocuadrático

Para evaluar la eficiencia de los estimadores de una superficie de respuesta se han propuesto varios procedimientos basados en el sesgo, como es el caso del error cuadrático medio, del cual sean derivado otros como el propuesto por Box y Draper en 1959, donde plantean el criterio del error cuadrático medio ponderado (criterio J), definido como:

$$J = \frac{N\Omega^{-1}}{\sigma^2} \int_R E[\hat{y}(x) - \eta(x)]^2 \quad (3)$$

Donde N es el número de corridas o puntos experimentales, $\Omega = \int_R dx$, R es la región de experimentación, σ^2 es la varianza del error experimental, $\hat{y}(x)$ es la respuesta estimada en el punto x , $\eta(x)$ es la respuesta verdadera en el punto x , $x' = x_1, x_2 \dots x_p$ donde p es el número de factores.

El criterio J puede expresarse como la suma de las varianzas y del sesgo integrado, de la respuesta estimada y tiene la propiedad de invarianza [17].

En este trabajo, el modelo utilizado parte de que la relación funcional desconocida $\eta = f(\xi_1, \xi_2, \dots, \xi_p)$ entre una variable de respuesta η y p variables explicativas donde $\xi_1, \xi_2, \dots, \xi_p$ se puede aproximar en alguna región de interés por una superficie de respuesta de segundo orden en términos de p variables codificadas $x_1 + x_2 + \dots + x_p$, las cuales son funciones lineales simples de las variables explicativas, esto es $y_i = f(x_{i1}, x_{i2}, \dots, x_{ip}) = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}^{aj} + \sum_{j=1}^p \beta_{jj} x_{ij}^{aj} + \sum_{j=1}^p \sum_{j < k} \beta_{jk} x_{ij}^{aj} x_{ik}^{ak} + \varepsilon_i, 1 = 1, \dots, n$

Donde las β son parámetros por estimar, x_{ij} es el nivel del factor j para la observación i , y_i es la respuesta de la observación experimental, ε_i es el error de ajuste del modelo correspondiente a las observaciones.

Una notación más simple y usual para un modelo de primer orden lineal sin interacciones o productos cruzados:

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \varepsilon \quad (4)$$

El modelo lineal de primer orden con interacciones

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=2}^k \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \varepsilon \quad (5)$$

Y el modelo cuadrático de segundo orden

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{j=2}^k \sum_{i=1}^{j-1} \beta_{ij} x_i x_j + \sum_{i=1}^{j-k1} \beta_{ij} x_{ii} x^2 + \varepsilon \quad (6)$$

Donde ε representa el ruido o valor observado en la respuesta y .

En la metodología de superficie de respuesta se parte del supuesto de que la variable respuesta y está en función de los niveles de los factores cuantitativos representados por las variables x_1, x_2, \dots, x_k .

Los modelos polinomiales se utilizan como una aproximación a la función de respuesta real, y generalmente son buenas aproximaciones cuando se trabaja en pequeñas zonas de los factores cuantitativos.

Cuando se trabaja con dos factores y se utiliza el modelo lineal ajustado de primer orden $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, la superficie de respuesta y sus curvas de nivel son las líneas con valores iguales de la respuesta, pero si el modelo es de segundo orden, su modelo ajustado es

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{11} x_1^2 + \hat{\beta}_{22} x_2^2 + \hat{\beta}_{12} x_1 x_2$ donde las superficies de respuesta y las gráficas de contornos son conocidas como curvas de nivel, las cuales pueden tener representaciones de un mínimo, una cresta elevada, una silla de montar, etcétera.

En la fase inicial del estudio de una superficie de respuesta se trata de identificar la región de respuesta óptima donde se utilizan experimentos factoriales completos 2^k o fraccionarios 2^{k-p} con el fin de estimar las respuestas medias para un modelo lineal o de primer orden, como. Generalmente se agregan dos o más observaciones en el nivel medio de cada uno de los factores para estimar el error experimental y tener un mecanismo de evaluación para saber si el modelo lineal es apropiado. Identificada la región de respuesta óptima, si los diseños factoriales completos o fraccionarios a dos niveles no son suficientes, se requieren al menos tres niveles para cada factor y el diseño debe de tener $1 + 2k + k(k - 1)/2$ puntos distintos para estimar los parámetros de un modelo de regresión cuadrática. Sin embargo utilizar factoriales 3^k necesita un número de combinaciones de tratamientos poco práctico, como es el caso de $k = 2$ factores, que necesita 9 combinaciones de tratamientos y agregar un factor más, esto es, tener un diseño factorial 3^3 requiere 27 combinaciones de tratamientos. Existen varias clases de diseños desarrollados para la aproximación a una superficie de segundo orden, que no demandan tantas combinaciones de tratamientos como los diseños factoriales 3^k y donde cada uno de ellos posee ciertas características y propiedades. Entre estos están los diseños centrales compuestos propuestos por Box y Wilson en 1951, que no crecen tanto como los diseños factoriales 3^k y los diseños Box-Behnken. Depende de los parámetros y el problema en particular la elección del modelo y para este caso multiobjetivo, Box-Behnken consiguió un mejor ajuste que un diseño compuesto central [15, 18]

VI. APLICACIÓN DE BOX BHENKEN Y SUPERFICIE DE RESPUESTA PARA UN PROBLEMA MULTI OBJETIVO

Entendido que el diseño experimental utiliza técnicas como la regresión múltiple, la superficie de respuesta, el

diseño de parámetros y varias extensiones del análisis de varianza, así como los análisis gráficos y las comparaciones múltiples, es posible asegurar que la aplicación de un diseño experimental (DE) aporta suficiente evidencia estadística para resolver la calibración de parámetros de un problema multiobjetivo como el que nos ocupa-

Diseñar un experimento significa planear un experimento de modo que reúna la información pertinente al problema bajo investigación. El DE es la secuencia completa de pasos como los que hemos señalado para asegurar que los datos se obtendrán apropiadamente de modo que permitan un análisis objetivo que conduzca a deducciones válidas con respecto al problema bajo estudio.

Para el problema multiobjetivo que optimiza compacidad geométrica y homogeneidad para variables censales, el objetivo es encontrar un balance de los parámetros de VNS (búsqueda local LS y estructuras de vecindad NS) para las 2 medidas de calidad (compacidad COM, homogeneidad HOM). Estas dos funciones objetivo se entiende como los factores que se estiman e impactan en la calibración de la obtención del mejor óptimo en la superficie de respuesta. Para conseguir esta meta, se debe estimar el dominio de interés de los parámetros y cómo se miden.

Usando la información previamente obtenida de un balance para LS y NS en el caso de un sólo objetivo, se pudo definir el área de experimentación para este trabajo [1, 2]. En este aspecto, nos esforzamos para encontrar el área de experimentación comenzando con valores altos para NS y LS de acuerdo a un objetivo e ir bajando los valores hasta que el costo computacional no rebasara 1 minuto. Así definimos el área de experimentación como se muestra en la siguiente tabla 2, donde solo valores enteros son válidos:

- a) $2 \leq NS \leq 6$,
- b) $10 \leq LS \leq 30$ y agregamos la el parámetro grupos (G)
- c) $8 \leq G \leq 24$.

Las 75 diferentes combinaciones que se muestran en la Tabla II exigieron ser probadas con nuestro programa multiobjetivo. Una vez realizadas las pruebas para cada combinación de los parámetros, las 2 últimas columnas se llenaron con los resultados arrojados de las corridas correspondientes.

Aquí es importante aclarar que el diseño desarrollado fue desbalanceado pero lo adaptamos al caso balanceado. Un trabajo para caso desbalanceado merece otro apartado mayor.

TABLA II
COMBINACIÓN DE PARÁMETROS PARA 75 PRUEBAS

	Pruebas factoriales					
	<i>Bloque</i>	<i>NS</i>	<i>LS</i>	<i>G</i>	<i>COM</i>	<i>HOM</i>
1	2	2	10	16	1792547	219309
2	2	6	10	16	1268552	306579
3	2	2	30	16	1463123	244264
4	2	6	30	16	1215842	351145
5	2	2	20	8	2139598	200530
6	2	6	20	8	2025129	175853
7	2	2	20	24	972303	381661
8	2	6	20	24	924503	423923
9	2	4	10	8	2070166	242748
10	2	4	30	8	2013185	204081

	Pruebas factoriales					
	<i>Bloque</i>	<i>NS</i>	<i>LS</i>	<i>G</i>	<i>COM</i>	<i>HOM</i>
11	2	4	10	24	937042	382943
12	2	4	30	24	1117403	300831
13	0	4	20	16	1221065	557653
14	0	4	20	16	1493049	247242
15	0	4	20	16	1296191	370661
16	2	2	10	16	1648394	263425
17	2	6	10	16	1260483	341323
18	2	2	30	16	1196954	477997
19	2	6	30	16	1577878	234653
20	2	2	20	8	2113342	209254
21	2	6	20	8	2303055	60811
22	2	2	20	24	1060539	338129
23	2	6	20	24	1320765	297713
24	2	4	10	8	2085672	158199
25	2	4	30	8	2044510	196822
26	2	4	10	24	982245	292835
27	2	4	30	24	973855	330943
28	0	4	20	16	1269926	288779
29	0	4	20	16	1261276	261743
30	0	4	20	16	1304619	323257
31	2	2	10	16	1236188	562793
32	2	6	10	16	1422914	245019
33	2	2	30	16	1352142	258048
34	2	6	30	16	1246257	267487
35	2	2	20	8	2066627	249804
36	2	6	20	8	2152002	144067
37	2	2	20	24	937581	395146
38	2	6	20	24	939075	302531
39	2	4	10	8	2157442	128013
40	2	4	30	8	2307670	131230
41	2	4	10	24	1229372	233831
42	2	4	30	24	1050689	322345
43	0	4	20	16	1255220	424799
44	0	4	20	16	1533651	213559
45	0	4	20	16	1258626	379346
46	2	2	10	16	1247852	301595
47	2	6	10	16	1441629	244307
48	2	2	30	16	1212100	326335
49	2	6	30	16	1416207	239937
50	2	2	20	8	1866926	414813
51	2	6	20	8	2229331	93620
52	2	2	20	24	1054690	349267
53	2	6	20	24	923389	526971
54	2	4	10	8	1859169	445411
55	2	4	30	8	1899160	352465
56	2	4	10	24	1166429	289517
57	2	4	30	24	1287079	289089
58	0	4	20	16	1258404	357549
59	0	4	20	16	1234663	356782
60	0	4	20	16	1690021	221707
61	2	2	10	16	1654590	263173
62	2	6	10	16	1180549	541436
63	2	2	30	16	1870944	222729
64	2	6	30	16	1405096	250235
65	2	2	20	8	2172206	189847
66	2	6	20	8	1786126	350599

	Pruebas factoriales					
	Bloque	NS	LS	G	COM	HOM
67	2	2	20	24	1125316	329902
68	2	6	20	24	1154594	298235
69	2	4	10	8	1864318	387121
70	2	4	30	8	2271781	163544
71	2	4	10	24	896542	384137
72	2	4	30	24	905142	540563
73	0	4	20	16	1326614	254733
74	0	4	20	16	1220163	465182
75	0	4	20	16	1306829	318929

Con los valores de la Tabla II, se desarrolló un modelo de experimentación tipo Box Bhenken con 5 réplicas. A partir de este modelo, se tiene un modelo de segundo orden para compacidad COM, el cual se presenta en la tabla III con los cálculos asociados. Se denota Coef para los coeficientes del modelo de la ecuación de regresión, SE Coef representa el error estándar del coeficiente, T es la prueba T-Student y P el coeficiente de correlación de Pearson. En el análisis de varianza se observa un coeficiente de correlación de .93 y un coeficiente de falta de ajuste P=.433, lo cual indica un buen ajuste estadístico [5, 15].

TABLA III
COEFICIENTES ESTADÍSTICOS DEL MODELO PARA COMPACIDAD

Término	Coef	SE Coef	T	P
constante	3531791	368909	9.574	0
NS	-133939	94066	-1.424	0.159
LS	-16765	18813	-0.891	0.376
G	-155214	23516	-6.6	0
NS*NS	10529	9631	1.093	0.278
LS*LS	347	385	0.901	0.371
G*G	3009	602	4.998	0
NS*LS	1929	1851	1.042	0.301
NS*G	-78	2313	-0.034	0.973
LS*G	-236	463	-0.509	0.612

S = 165534, R-cuad. = 86.33% R-cuad.(pred.) = 81.54% R-cuad.(ajustado) = 84.43%.

La tabla III, puede reescribirse en una ecuación como modelo de segundo orden si denotamos $YC = compacidad$, $X_1 = NS$, $X_2 = LS$, $X_3 = G$:

$$YC = 3531791 - 13393X_1 - 16765X_2 - 15521X_3 + 10529X_1^2 + 347X_2^2 + 3009X_3^2 + 1929X_2X_3 - 78X_1X_3 - 236X_2X_3 \quad (7)$$

Se observa en Fig. 4 las pruebas de validez de los resultados del modelo para compacidad. Se concluye que los datos se comportan normalmente, que el modelo de segundo orden es adecuado y que no existen efectos de sesgo de una corrida a otra en el experimento. Estos resultados aseguran que es posible continuar con el experimento.

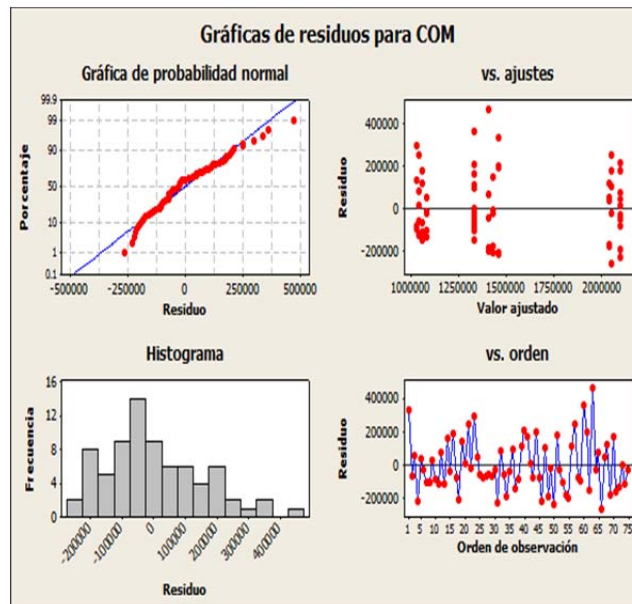


Figura 4. Gráfico que comprueban los supuestos de modelo para compacidad (COM).

Para la respuesta de homogeneidad, el modelo encontrado viene dado como sigue:

TABLA IV
COEFICIENTES ESTADÍSTICOS DEL MODELO PARA HOMOGENEIDAD

Término	Coef	SE Coef	T	P
Constante	163244	219028	0.745	0.459
NS	11954	55849	0.214	0.831
LS	1250	11170	0.112	0.911
G	12431	13962	0.890	0.377
NS*NS	-3780	5718	-0.661	0.511
LS*LS	-129	229	-0.565	0.574
G*G	-537	357	-1.503	0.138
NS*LS	-636	1099	-0.579	0.565
NS*G	1546	1374	1.125	0.265
LS*G	321	275	1.169	0.247

S = 98280.4, R-cuad. = 26.47% R-cuad.(pred.) = 1.78% R-cuad.(ajustado) = 16.29%.

En este caso, el análisis de varianza arrojó un coeficiente de correlación de .51 y un valor P de falta de ajuste de .728. [5, 15]. La tabla IV se reduce a una ecuación sencilla si denotamos $YH = homogeneidad$, $X_1 = NS$, $X_2 = LS$, $X_3 = G$. Entonces, la ecuación asociada se expresa como sigue

$$YH = 163244 + 11954X_1 + 1250X_2 + 12431X_3 - 3780X_1^2 - 129X_2^2 - 537X_3^2 - 636X_1X_2 + 1546X_1X_3 + 321X_2X_3 \quad (8)$$

A partir de esta información se han generado los siguientes gráficos que permiten tener mayor información sobre el objetivo de balanceo del problema. Note en la Fig. 5 la verificación del modelo. Es evidente la distribución normal para los datos con un modelo de segundo orden aceptable y además los efectos de una corrida a otra no alteran el experimento.

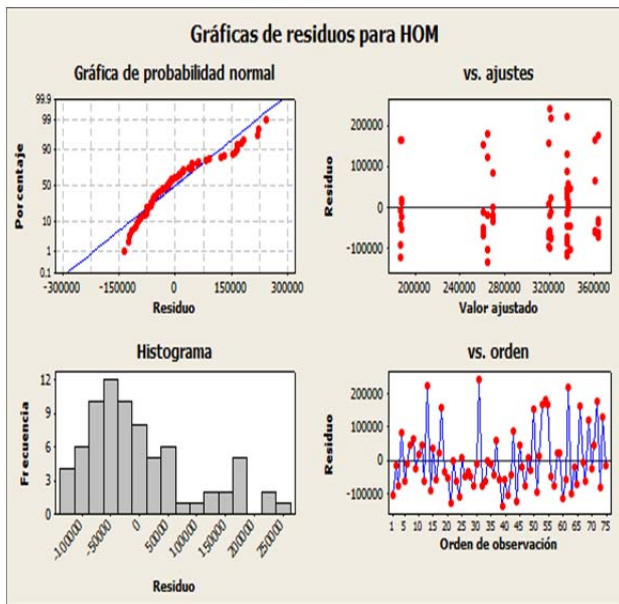


Figura 5. Gráfico que comprueba los supuestos del modelo para homogeneidad (HOM).

A. Análisis de las gráficas de contorno

Las gráficas de contorno son útiles para identificar visualmente las variables factibles para múltiples respuestas en un experimento de diseño factorial, de superficie de respuesta o de mezcla, tal como el que hemos desarrollado en este trabajo. La configuración de variables factibles para una respuesta podría estar lejos de ser factible para otra respuesta y las gráficas superpuestas de contornos pueden usarse para considerar las respuestas simultáneamente.

Cuando se crea una gráfica superpuesta de contornos, esta especifica un límite inferior y superior para cada respuesta. La gráfica superpuesta de contornos muestra contornos para esos límites con respecto a dos factores continuos. Las demás variables del modelo se mantienen en los valores de configuración especificados. Una línea de contorno es una curva que conecta los puntos de la gráfica de manera que los valores de respuesta ajustados sean iguales. Por ejemplo, con un límite inferior de 30, el contorno para el límite inferior sería una curva que conecte los puntos de la gráfica con los valores de respuesta ajustados iguales a 30. La gráfica destaca la región (si existe) donde todas las respuestas están dentro de sus límites. Esto se conoce como la región factible.

En este apartado se muestran algunas gráficas de contorno que son generadas con conclusiones que obedecen al análisis de superficies de respuesta. Convenimos expresarlas en contornos para una mejor interpretación. Esto es, al identificarse en que regiones se alcanzan valores cercanos al óptimo de la función objetivo, se han graficado contornos de comportamiento de esta función para regiones donde los parámetros de predicción de la misma son los adecuados.

El propósito de interpretar las gráficas de contorno de la compacidad independiente de la homogeneidad constituye uno de los resultados más importantes de este trabajo: encontrar la calibración de homogeneidad y compacidad cuando actúan de forma simultánea, es decir, una vez identificados los parámetros en balance sobre la zona de experimentación donde el óptimo es alcanzado para cada objetivo, ha sido

posible combinarlos para establecer la calibración adecuada de los parámetros para los dos objetivos de interés. Las siguientes subsecciones se encargan de dar una explicación concreta.

1) Compacidad: Análisis de zona de experimentación bajo gráficas de contorno para 24 grupos

Hemos elegido el caso para 24 grupos de acuerdo a la combinación de parámetros de la Tabla II.

En la siguiente Fig. 6 es posible observar las diferentes curvas de nivel y la zona donde se alcanza el mínimo-óptimo para compacidad alrededor de 102000 con LS en 11 y NS entre 5 y 6.

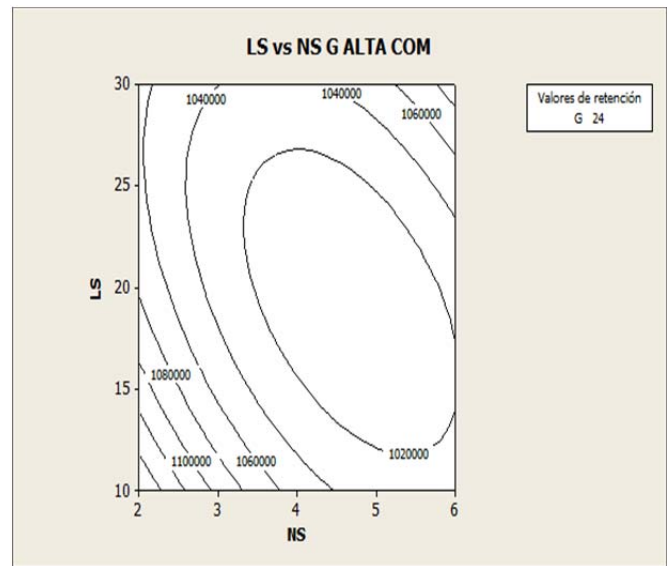


Figura 6. Curvas de nivel para compacidad con 24 grupos.

2) Homogeneidad: Análisis de zona de experimentación bajo gráficas de contorno para 24 grupos

Fig. 7 presenta las curvas de nivel donde se identifica el contorno bien definido para alcanzar mínimo, el cual se encuentra en 320000, con LS en 11 y NS entre 2 y 3.

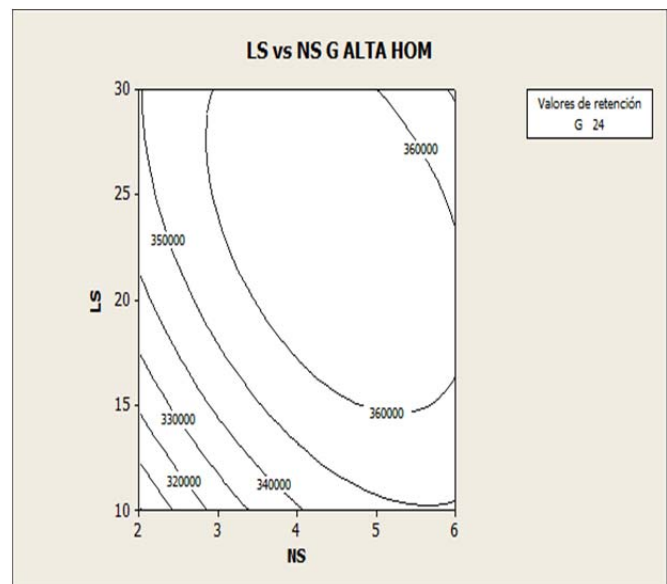


Figura 7. Curvas de nivel para homogeneidad con 24 grupos.

3) Zona óptima de experimentación para compacidad-homogeneidad para 24 grupos

Para conseguir que NS y LS generen valores mínimos simultáneamente de compacidad y homogeneidad para 24 grupos, se tuvo que restringir a las siguientes regiones de acuerdo al análisis de las gráficas de Fig. 5 y Fig. 6.

En Fig. 8, la gráfica muestra una clara intersección de la zona de experimentación para alcanzar el óptimo (mínimo) para homogeneidad y compacidad. La zona blanca es tal intersección.

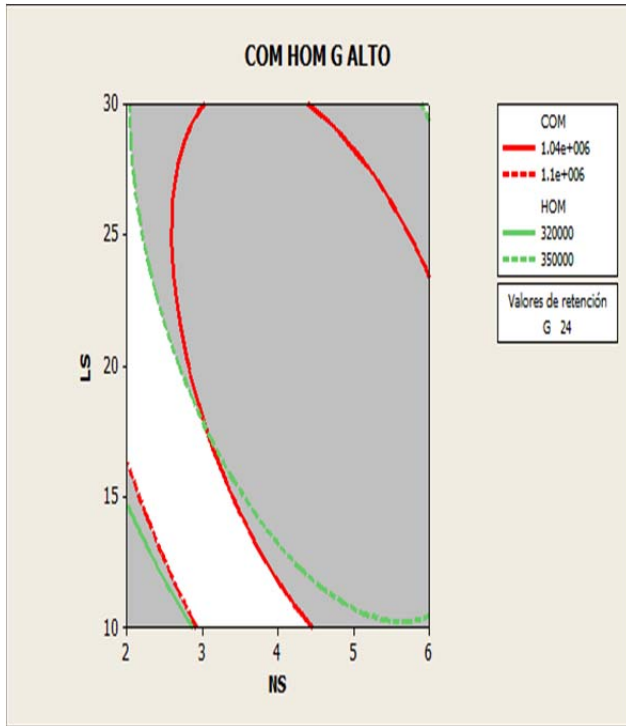


Figura 8. Gráfico de contornos para HOM, COM con 24 grupos.

La zona “blanca” que se aprecia en Fig. 8 es una región en donde todos los puntos que se encuentran dentro corresponden a una combinación de parejas NS y LS que satisfacen simultáneamente la condición de minimización entre compacidad y homogeneidad. Esta región está delimitada por 4 puntos:

TABLA V
COMBINACIÓN DE PARÁMETROS PARA HOMGENEIDAD-
COMPACIDAD (24 GRUPOS)

Contorno para HOM-COM, 24 grupos	
LS	NS
29.9	2.0
16.3	2.0
10.0	2.9
10.0	4.4

Los valores indicados en la tabla son la respuesta adecuada que conducen a que homogeneidad y compacidad alcancen el mínimo cuando ambos parámetros se optimizan de forma simultánea.

B. Análisis de las gráficas de contorno para 16 grupos

En esta subsección, se desarrolló el cálculo para el caso de 16 grupos. El procedimiento es análogo a 24 grupos donde se obtuvo la gráfica para el comportamiento estimado de homogeneidad para proceder de la misma forma con la compacidad. Los siguientes gráficos permiten tener mayor información sobre el objetivo de balanceo que hemos planteado.

1) Homogeneidad: Análisis de zona de experimentación bajo gráficas de contorno.

La siguiente Fig. 9 se ha analizado del mismo modo que se realizó para 24 grupos. Existe una zona óptima lejos del mínimo que puede verse al centro de la figura con un valor de 34000. El mínimo se aprecia cerca de 30000 con LS 30 y NS entre 5 y 6.

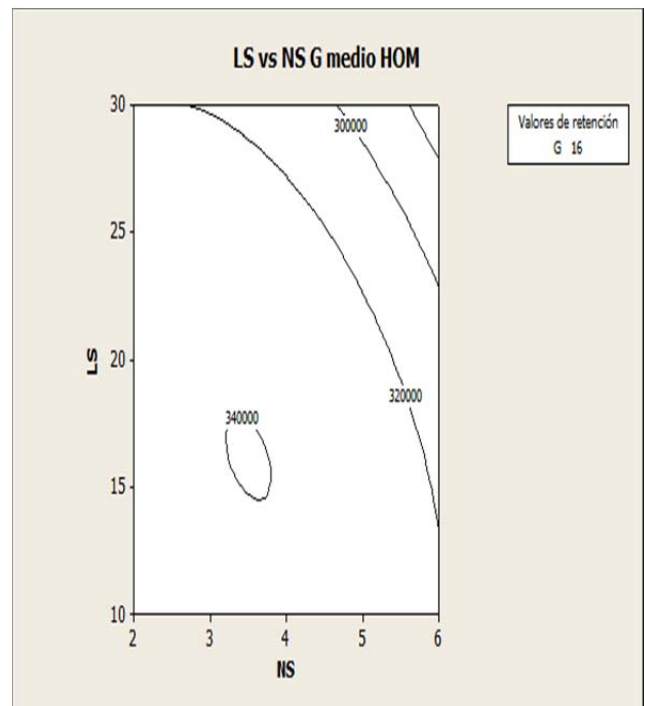


Figura 9. Curvas de nivel para homogeneidad con 16 grupos.

2) Compacidad: Análisis de zona de experimentación bajo gráficas de contorno.

La siguiente gráfica Fig. 10, revela la zona de experimentación para compacidad con 16 grupos. La zona donde se concentran los mínimos se encuentra entre 134000 con NS entre 4 y 5 y LS en 11.

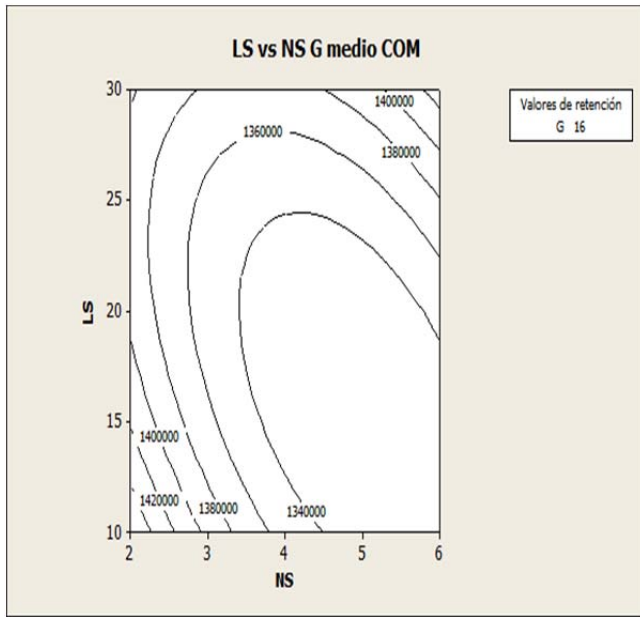


Figura 10. Curvas de nivel para compacidad con 16 grupos.

3) Zona óptima de experimentación para compacidad-homogeneidad para 24 grupos.

La siguiente gráfica Fig. 11 muestra las condiciones de balanceo para alcanzar el mínimo de COM y HOM para 16 grupos.

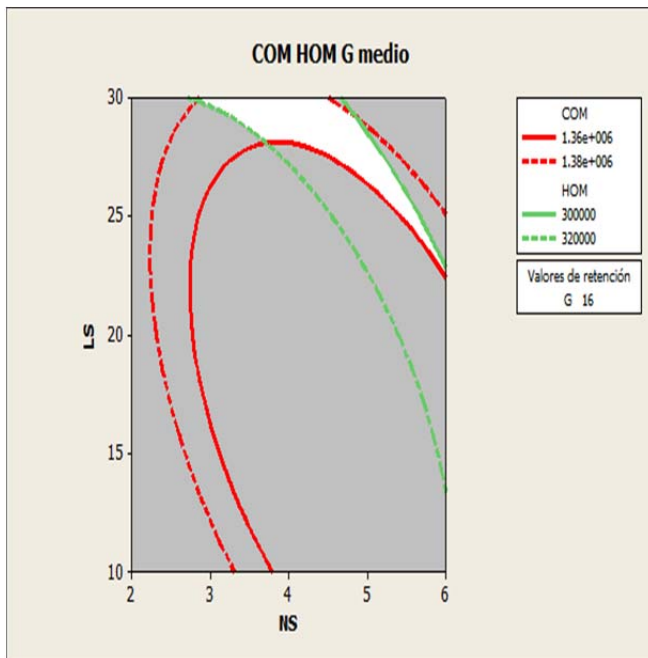


Figura 11. Gráfico de con tornos para CO,M HOM con 16 grupos

TABLA VI
COMBINACIÓN DE PARÁMETROS PARA HOMOGENEIDAD-
COMPACIDAD (16 GRUPOS)

Contorno para HOM-COM, 16 grupos	
LS	NS
29.91	2.82
29.94	4.55
28.15	3.68

Contorno para HOM-COM, 16 grupos	
LS	NS
22.87	6.0
22.43	6.0

La Tabla VI indica los valores adecuados que permiten que homogeneidad y compacidad alcancen el mínimo cuando ambos parámetros se optimizan de forma simultánea.

VII. CONCLUSIONES

A lo largo de este trabajo hemos podido concentrar diversos resultados. El primero se centra en haber conseguido una agrupación multiobjetivo para un problema poblacional sobre población económicamente inactiva. Este tipo de problemas está relacionado con proyectos políticos donde se asume existen recursos para atender población vulnerable económicamente. Entonces con una agrupación como la obtenida, es posible identificar que grupos merecen más atención. El mapa debe analizarse cuidadosamente con un SIG para revisar la concentración de la población con las características que hemos señalado, sin embargo este análisis debe ser tratado en un artículo independiente.

Finalmente es necesario subrayar la contribución del presente artículo: mientras otros métodos encuentran el FP con un método externo donde toman en la entrada todas las soluciones que se generan en el proceso de agrupamiento, nuestro algoritmo encuentra las soluciones dominadas dentro del proceso de agrupamiento con VNS. Esto es, nuestro programa multiobjetivo recibe 2 matrices en la entrada, una es de distancias y otra de variables poblacionales. La salida es el FP. Por otra parte, no es fácil calibrar los parámetros de VNS y más aún cuando de problemas multiobjetivo se trata. Otra aportación se centra en el diseño de un experimento factorial para encontrar un balance en los parámetros y determinar la mejor combinación que alcance un conjunto de soluciones en tiempo de cómputo razonable y calidad en las soluciones.

A pesar de que la agrupación multiobjetivo fue alcanzada con valores aleatorios para los parámetros NS y LS, la calibración de estos fue uno de los principales desafíos de este trabajo. La experiencia obtenida al calibrar parámetros de la metaheurística VNS para un objetivo, contribuyó a explorar y aplicar un experimento factorial para 2 objetivos y así establecer un balanceo objetivo de los parámetros.

Una vez que obtuvimos la calibración de los parámetros para compacidad y homogeneidad, observamos que el tiempo de respuesta no rebasa el minuto. Este incremento en tiempo lo esperábamos debido a la presencia de dos objetivos. Sin embargo ha sido satisfactoria la combinación de parámetros conseguida con la metodología de superficie de respuesta, la cual ha sido poco explorada y aplicada para encontrar balance de los parámetros en metaheurísticas.

Dado que en este trabajo trabajamos un caso balanceado de diseño de experimentos, el reto es desarrollar el experimento con desbalanceo.

REFERENCIAS

- [1] B. Bernábe, J. Ramírez, J. Espinosa. Evaluación de un algoritmo de recocido simulado con superficies de respuestas, Revista Matemática Teoría y Aplicaciones, vol. 6 (1), 2009, pp 159-177.

- [2] B. Bernábe, J. Ramírez, J. Espinosa, M. A. Osorio, R. Aceves, An Adjusted Variable Neighborhood Search Algorithm applied to the Geographical Clustering Problem Research in Computer Science issn 1870-4069, vol. 42, 2009, pp. 113-125.
- [3] M. Altman, The Computational Complexity of Automated Redistricting: Is Automation the Answer?, Rutgers Computer and Technology Law Journal, vol. 23 (1), 1997, pp. 81-142.
- [4] H. T. Kung, F. Luccio, & F. P. Preparata, On Finding the Maxima of a Set of Vectors, Journal of the ACM (JACM), vol 22(4), 1975, pp. 469-476.
- [5] D. Montgomery. Design and Analysis of Experiments. Wiley, 2 edition, 1991-
- [6] R. Caballero, T. Gomez, M. Gonzalez, M. Muñoz, M. Rey, F. Ruiz, Programación Matemática para Economistas, Universidad de Málaga, ISBN 84-7496-651-5, 1997, Málaga.
- [7] L. A. Lara, Un estudio de las Estrategias Evolutivas para problemas Multi-objetivo, Tesis de Maestría en Ciencias en la especialidad de Ingeniería Eléctrica Opción Computación, Cinvestav IPN, , 2003, pp. 12-29.
- [8] M. B. Loranca & C. G. Galván, Búsqueda de entorno variable multiobjetivo para resolver el problema de particionamiento de datos espaciales con características poblacionales. Computación y Sistemas, vol. 16 (3), 2012, pp. 335-347.
- [9] INEGI. (2000). Sistema para la consulta de información censal 2000, (SCINCE), XII Censo General de Población y Vivienda 2000. Retrieved from=http://www.inegi.org.mx/prod_serv/contenidos/espanol/catalogo/Default.asp?accion=2&upc=702825496371
- [10] P. Hansen, N. Mladenovic, M. Pérez, Variable Neighbourhood Search., Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial. Number 19, 2003, pp. 77-92.
- [11] D. Ríos, Sobre soluciones optimas en problemas de optimización multiobjetivo, Trabajos de Investigación Operativa. Editor Springer Berlin/ Heidelberg, vol. 2 (1) , pp. 2008, 49-67.
- [12] B. Bernábe, Desarrollo de un Modelo para la Determinación de Zonificación Óptima, PhD thesis, División de Estudios de Posgrado de la Facultad de Ingeniería DIMEI, UNAM, Investigación de Operaciones, 2006, pp. 73-91.
- [13] NODOM
<http://www.cs.cinvestav.mx/~emoobook/nodom/nonodom.html>
- [14] B. Bernábe, R. Gonzalez, Integración de un sistema de información geográfica para algoritmos de particionamiento, Research in Computing Science, vol. 88, 2014, pp. 31-44.
- [15] G. E. Box, K. G. Wilson, On the experimental attainment of optimum conditions, Journal of the Royal Statistical Society, vol 13, 1951, pp. 1-45
- [16] F. B. Cady and R. J. Laird. Treatment design for fertilizer use experimentation, CIMMYT Research Bulletin, vol. 26, 1973, pp. 29-30.
- [17] G. E. Box and N. R. Draper, A basis for the selection of a response surface design, J. Amer Statist. Assoc., vol. 54, 1959, pp. 622-654.
- [18] G. F. Preciado Mosaicos, Optimización de una superficie de respuesta utilizando JMP IN, Mosaicos Matemáticos No. 11 Nivel Superior, Departamento de Matemáticas Universidad de Sonora , 2003., pp.17-23.
- [19] <http://www.monografias.com/trabajos-pdf5/multiobjective-tabu-search-for-the-zoning-problem/multiobjective-tabu-search-for-the-zoning-problem.shtml>



María Beatriz Bernábe Loranca es Doctora en Investigación de Operaciones con mención honorífica por la UNAM (2010). Maestría en Ingeniería en Calidad por la Universidad IBEROAMERICANA en el año 2003 y Licenciada en Ciencias de la Computación por la Facultad de Ciencias Físico-Matemáticas por la BUAP 1994.

Realizó una estancia posdoctoral entre 2011-2013 con apoyo Conacyt en la UPAEP en el posgrado de Logística y Dirección de la Cadena de Suministro.

Actualmente es Profesora-Investigadora de Tiempo Completo Titular A en la Facultad de Ciencias de la Computación donde imparte Matemáticas Discretas, Probabilidad, Estadística, Investigación de Operaciones, Lenguajes Formales y Autómatas y Simulación.

Dentro de la investigación, las áreas de interés son Optimización Combinatoria, Diseño Territorial, Optimización Multiobjetivo, Logística y Metaheurísticas.

Actualmente pertenece al Sistema Nacional de Investigadores nivel I (2016-2018)



Jorge A. Ruiz-Vanoye obtuvo el grado de Doctor en Ciencias Computacionales en 2008 por el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET). Ha trabajado en el Instituto de Investigaciones Eléctricas y en otras instituciones y compañías. Ha dado clases en diversas universidades mexicanas desde 1996. Actualmente es profesor investigador en la Universidad Autónoma del Estado de Hidalgo (UAEH) y es miembro del Sistema Nacional de Investigadores, nivel I (2013-2019)



Mario Mauricio Bustillo Díaz obtuvo el título de ingeniero en el Instituto Politécnico de Odesa (Ucrania- CCCP) en Ciencias de la Computación y la Maestría en Ciencias en el área de Arquitecturas de Computadoras. Recibió el título de Ph.Dr. en la Universidad Politécnica Estatal de San Petersburgo (Rusia), en la especialidad de sistemas difusos. Actualmente se desempeña como Profesor-Investigador en la Benemérita Universidad Autónoma de Puebla, México.



Rogelio González Velázquez obtuvo el grado de doctor en Logística y Cadena de Suministro en la Universidad Popular Autónoma del Estado de Puebla, el grado de maestro en Investigación de Operaciones en la Universidad Nacional Autónoma de México, es desde 1985 profesor investigador en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla, en Puebla México, desarrolla la línea de investigación y generación del conocimiento de investigaciones en el cuerpo académico de computación matemática y es integrante del padrón de investigadores de la BUAP.



Alberto Ochoa Ortiz - Zezzatti (Licenciatura en Informática'94 – Maestría en Ingeniería'00; Doctorado en Tecnología Avanzada'04 – Postdoctorado en Investigación'06 y Postdoctorado en la Industria'09). Se incorporó a la Universidad Autónoma de Ciudad Juárez en 2008. Participa en la organización de varias conferencias internacionales. Sus intereses de investigación incluyen la computación ubicua, la computación evolutiva, el procesamiento del lenguaje natural, modelado social inteligente, caracterización antropométrica y Minería de datos social. En mayo de 2016 comienzan un año sabático de investigación en Barcelona Supercomputing Center con un proyecto relacionado con la simulación de estampidas humanas y el hundimiento de barcos de refugiados mejorado mediante la Inteligencia Artificial.



Gerardo Martínez Guzmán obtuvo el título de Licenciado en Matemáticas. Maestría y Doctorado en Ciencias Matemáticas por la Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla, México. Actualmente es profesor investigador de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla, México.