

# 2<sup>do</sup> Taller LATino Iberoamericano de Investigación de Operaciones

## “La IO aplicada a la solución de problemas regionales”

### Conglomerados y Heurísticas en Problemas de Regionalización

María Beatriz Bernábe L<sup>1</sup>., Ma. Auxilio Osorio L<sup>2</sup>, Ricardo Aceves G<sup>3</sup>, Alba Sánchez G<sup>4</sup>.

UNAM

<sup>1,3</sup> División de Estudios de Posgrado de la Facultad de Ingeniería  
Departamento de Sistemas-Investigación de Operaciones

<sup>2,4</sup> Benemérita Universidad Autónoma de Puebla  
Facultad de Ciencias de la Computación

**Resumen:** El problema de regionalización consiste en la agregación de áreas geográficas en un número predefinido de regiones homogéneas y espacialmente continuas. La solución del problema de regionalización es de carácter combinatorio (“NP-hard”) y exige que su solución sea generada a partir de métodos no exactos de optimización. Dentro de las aproximaciones metodológicas se observa la aplicación y readaptación de métodos convencionales de agregación (clasificación jerárquica o de partición) en los cuales la restricción de continuidad geográfica se satisface indirectamente a través de la inclusión de las variables geográficas (por ejemplo distancias entre áreas) dentro del grupo de variables de clasificación. En este trabajo, para dar respuesta a diversos problemas censales de regionalización, hemos implementado un algoritmo propio de particionamiento teniendo como principal objetivo el minimizar distancias entre objetos y centroides. Debido a que este algoritmo tiene un alto costo computacional, se hace necesario mejorar su desempeño en este punto. Para ello hemos incorporado el método heurístico conocido como recocido simulado dentro del algoritmo de conglomerados con el fin de dotar a dicho algoritmo de herramientas adicionales que le permitan escapar con mayor eficiencia de soluciones subóptimas, las cuales suelen ser muy comunes en procesos de regionalización. Para comprender el estado actual de este trabajo, presentamos breve recorrido histórico que precede a los resultados que hasta hoy se han logrado.

Keywords: agregación, conglomerados, particionamiento, recocido simulado, regionalización.

### Introducción

El diseño de territorio puede ser visto como un problema de agrupación de áreas geográficas pequeñas

(áreas básicas) en clusters geográficos más grandes llamados territorios, del tal forma que la agrupación aceptable es aquella última que cumpla con ciertos criterios predeterminados del problema que ocupa [7]. Dependiendo del contexto, estos criterios bien pueden ser de motivo económico (promedio de ventas potenciales, trabajo o número de vendedores) o tienen un fondo demográfico (número de habitantes, población votante). Es más, restricciones espaciales como continuidad y compacidad son muy demandadas. En este sentido, es posible destacar que regionalización trata de abordar diversos problemas de diseño territorial. Por otro lado, se debe tener presente que estos problemas tienen implícita una alta complejidad computacional y se hace necesario que su solución sea generada a partir de métodos aproximados de optimización. Las aproximaciones metodológicas más utilizadas en este ámbito se centran en la aplicación de métodos clásicos de clasificación incorporando algún método heurístico. No importando demasiado el método de agregación creado, debe considerarse la restricción de continuidad geográfica, la cual suele satisfacerse si las variables geográficas se encuentran disponibles y claras tanto en coordenadas geográficas como en definición (en este caso variables censales). La inclusión de este tipo de variables favorecerá la creación de regiones espacialmente compactas, lo cual se traduce en la satisfacción de la restricción de continuidad geográfica [7,11]

Es en este marco donde se presenta este trabajo: Para dar respuesta al problema de regionalización de carácter censal sobre AGEBS (Áreas Geoestadísticas Básicas) [17], hemos implementado un algoritmo propio de particionamiento (conglomerados) teniendo como principal objetivo el minimizar distancias entre objetos y centroides. Debido a que este algoritmo tiene un alto costo computacional, se hace necesario mejorar su desempeño en este punto, de esta manera se ha

implementado la heurística conocida como recocido simulado dentro del algoritmo de conglomerados que hemos diseñado y que actúa sobre variables geográficas, de esta forma hemos logrado obtener buenas soluciones. La evidencia empírica que aportaremos estará basada en los 4952 AGEBS de la Zona Metropolitana del Valle de Toluca (ZMVT). Las variables de clasificación estarán conformadas por 57 variables socioeconómicas disponibles para dichas áreas. El presente trabajo se encuentra organizado como sigue: Introducción, generalidades sobre problemas de regionalización, heurística de Recocido Simulado y su asociado algoritmo de conglomerados, Pruebas y Conclusiones.

### **Sección I: Generalidades Sobre los Problemas de Regionalización**

Entre la planificación y diseño territorial, conviene la utilización de información estadística a nivel regional y urbano (que cada vez está siendo más utilizada para llevar a cabo análisis y planeación de carácter poblacional, económico, distribución, muestreo, etc) [1,15, 17].

La agregación de información territorial es necesaria en los problemas de diseño territorial, la mayoría de las veces suele hacerse de manera particular debido a la escasez de métodos de regionalización lo suficientemente flexibles. Además, independientemente de la metodología de agregación territorial que se utilice, existe un riesgo implícito que se ha tratado en la literatura bajo el nombre del Problema de la Unidad Espacial Modificable, PUEM [4] y que consiste en la elevada sensibilidad de los resultados a la agregación de datos geográficos, lo cual puede traer consecuencias no deseables en el análisis.

Asumiendo como válidas diferentes alternativas para solucionar problemas de categorización de unidades geográficas, estas requieren de un común denominador: analizar las zonas en un contexto geográfico bien definido, es decir, estudiar zonas muy pequeñas o muy grandes dificulta el análisis y solución del problema que se este tratando. Básicamente, un inicio de estudio de la población vs zonas, debe apoyarse en el supuesto de que si se obtienen agrupaciones de zonas equilibradas y contiguas, esto contribuye a abordar el problema con mas precisión y facilidad [17]

En este contexto, los primeros intentos de particionamiento para los AGEBS, se apoyaron de las

herramientas de clasificación que ofrecen algunos paquetes estadísticos, sin embargo, los resultados de agrupamiento arrojados distan mucho de los principios básicos de agregación territorial [2,10]. Estas agrupaciones cumplen con criterios predeterminados por el paquete estadístico, sin embargo en los grupos resultantes se observó que las unidades geográficas dispersas violando el principio de compacidad. Por otro lado, las soluciones que ofrecen son solo óptimos locales [10]. Como consecuencia de este inconveniente hemos propuesto otras técnicas para obtener clasificaciones más precisas, tales como la obtención de grupos “compactos” y homogéneos [15]. Aún así, el problema combinatorio persiste y es necesario proponer modelos de optimización combinatoria que ayuden a que el proceso de clasificación de una buena solución sin que se sacrifique el tiempo de computo. En este sentido ya hemos resuelto la incorporación de Recocido Simulado con 2 variantes a 2 diferentes métodos de clasificación: 1) en PAM [1] y 2) en un algoritmo de partición (con el método jerárquico no se lograron buenos resultados [10]). Por otro lado intentamos resolver el problema de clasificación recurriendo a técnicas de Bases de Datos y un algoritmo de particionamiento adecuado a la extracción de datos [15].

Otros autores han abordado el problema de regionalización, distribución territorial, diseño de territorio, etc., desde diferentes ángulos, contándose incluso con modelos de optimización combinatoria [113]. Estas propuestas incluyen algoritmos de resolución pseu-óptima pero aún hay mucho por hacer para resolver las fuentes de complejidad (no linealidad, no convexidad, múltiples objetivos, no compacidad ni homogeneidad). Una de las definiciones más precisas sobre diseño territorial puede entenderse como un problema de agrupación sobre áreas geográficas pequeñas (áreas básicas) en clusters geográficos más grandes llamados territorios, del tal forma que la agrupación aceptable es aquella última que cumpla con ciertos criterios predeterminados del problema que ocupa. Dependiendo del contexto, estos criterios bien pueden ser de motivo económico (promedio de ventas potenciales, trabajo o número de vendedores) o tienen un fondo demográfico (número de habitantes, población votante). Es más, restricciones espaciales como continuidad y compacidad son muy demandadas [7,11]. En la literatura existente hemos encontrado trabajos que han servido de apoyo para el problema que

estamos tratando, en el que destaca una propuesta metodológica de regionalización para el diseño de unidades geográficas que están directamente relacionadas con el fenómeno analizado donde se considera la imposición de homogeneidad y contigüidad geográfica en el proceso de agregación territorial [4]. Este tipo de restricciones son las más demandadas y en este trabajo solo nos hemos ocupado de la compacidad geométrica. Otro enfoque sobre el problema de diseño o planeación de zonas o diseño de zonas geográficas (también conocido como la generalización de distribución o caso particular de este), ocurre cuando  $n$  unidades de áreas son agregadas en  $k$  zonas tal que un valor de función es optimizado, dependiendo de las restricciones en la topología de las zonas (por ejemplo conectividad interior) [5,12]. Cabe señalar que la terminología sobre unidades geográficas es en esta área versátil y no es difícil comprender por que existe confusión de los lectores en esta. Por ejemplo en lugar de región, en [5], los autores utilizan más el término de zona. Por otro lado, detallan un proceso que se basa en el principio básico: cada zona está compuesta de un conjunto de unidades territoriales elementales donde un mapa de distrito se forma dividiendo el conjunto de unidades elementales en zonas conexas sin inclusiones. Cuando un criterio múltiple como compacidad y homogeneidad es considerado, el problema de enumerar todas las soluciones eficaces para tal modelo se conoce como NP-difícil.

## Sección 2: Heurística de Recocido Simulado en un Algoritmo de Conglomerados para el Problema de Regionalización.

Debido a que los problemas de agregación territorial exigen métodos no exactos para su solución donde a su vez es necesario crear un algoritmo de conglomerados propio para el problema, y, con el fin de encontrar una solución lo suficientemente cercana a la óptima, el marco en el que se sitúa el presente trabajo es la implementación de la metaheurística de Recocido Simulado para optimizar la función objetivo (minimización de distancias) del método de clasificación que hemos diseñado, el cual, se ha aplicado a unidades geográficas. Este método de conglomerados ha sido el resultado de varias lecturas de algoritmos de particionamiento. Un algoritmo que fue determinante para iniciar el trabajo heurístico está apoyado en PAM [8,14].

**Recocido Simulado:** El Enfriamiento o Recocido Simulado es un algoritmo de búsqueda por entornos con un criterio probabilístico de aceptación de soluciones basado en Termodinámica Estadística [4]. Un modo de evitar que la búsqueda local finalice en óptimos locales, hecho que suele ocurrir con los algoritmos tradicionales de búsqueda local, es permitir que algunos movimientos sean hacia soluciones peores. Pero si la búsqueda está avanzando realmente hacia una buena solución, estos movimientos “de escape de óptimos locales” deben realizarse de un modo controlado [9].

En el caso del Enfriamiento Simulado (ES), esto se realiza controlando la frecuencia de los movimientos de escape mediante una función de probabilidad que hará disminuir la probabilidad de estos movimientos hacia soluciones peores conforme avanza la búsqueda (y por tanto estamos más cerca, previsiblemente, del óptimo local).

El algoritmo básico de recocido simulado se describe como sigue:

```

INPUT (To,  $\infty$ , L(t), Tf)
T  $\leftarrow$  To Valor inicial del parámetro de control
Sact  $\leftarrow$  Genera solución inicial
WHILE T  $\geq$  Tf DO Condición de parada
  BEGIN
    FOR cont  $\leftarrow$  1 TO L(T) DO Velocidad de Enfriamiento (T)
      BEGIN
        Scand  $\leftarrow$  Selecciona solución_N(Sact)
        Generación de una nueva solución
         $\delta \leftarrow$  coste(Scand)-coste(Sact) Cálculo de la diferencia de costos
        IF U(0,1)  $<$   $e^{(-\delta/T)}$  OR  $\delta <$  0 Aplicación del criterio de aceptación
          THEN
            Sact  $\leftarrow$  Scand
          END
        T  $\leftarrow$   $\infty(T)$  Mecanismo de enfriamiento
      END
    END
  }Escribe como solución la mejor de las Sact visitadas}

```

**Algoritmo de Conglomerados:** Esta sección presenta el algoritmo de Recocido Simulado que minimiza la distancia de los objetos a sus centroides más cercanos, dentro de un método de conglomerados inspirado en PAM (Kaufman and Rousseeuw ,1987), aplicado a



unidades geográficas de la zona metropolitana del Valle de Toluca.

### Algoritmo

1.-

#### Inicio

Sea  $O = \{O_1, O_2, \dots, O_n\}$  Objetos a clasificar  
Sea  $M = \{M_1, M_2, \dots, M_k\}$  Una solución de  $K$  centroides  
 $T_0$  temperatura inicial  
 $T_f$  temperatura final  
 $L(t)$  Número de iteraciones que se van a realizar con la misma temperatura

#### OBTIENE SOLUCIÓN INICIAL

Construir centroides iniciales  $M_1, M_2 \dots M_k$   
Donde  $M_1$  es el objeto con el valor mas pequeño de

$$\sum_{i=1}^n d(i, m_1)$$

$M_2 > M_1$  con respecto a  $\sum_{i=1}^n \min_{t=1, \dots, k} d(i, m_t)$  ( $M_2$  es el siguiente valor mínimo a  $M_1$ ) ...

$$M_k > M_{k-1} \text{ con } \sum_{i=1}^n \min_{t=1, \dots, k} d(i, m_t)$$

#### 2.- Intercambio

Construye otra solución para que los demás objetos sean centroides

Repetir hasta lograr convergencia

Χονσιδεραρ τοδος λος παρεσ δε οβφετος (ι,φ) ταλ  
θυε

$$ι \in \{1, 2, \dots, n\} \quad φ \in \{1, 2, \dots, k\}$$

Y realizar el intercambio  $i \leftrightarrow j$  para cualquiera que disminuya más el objetivo.

$i \leftarrow 1$  posición del centroide que se va a cambiar

$j \leftarrow 1$  posición del objeto a variar

Calcular costos

$\text{CostoA} \leftarrow \text{Costo de la función objetivo (M)}$

Mientras  $T \geq T_f$

Para  $\text{cont} = 1$  hasta  $L(t)$  hacer:

*Perturbar la solución actual: cambiar objeto por centroide*

$C \leftarrow M$

$C_i \leftarrow C_j$  perturba

objetivo (C)  $\text{CostoC} \leftarrow \text{Costo de la función}$

$\delta \leftarrow \text{CostoC} - \text{CostoA}$

Si  $U(0,1) < e^{-\delta/T}$  o  $\delta < 0$  hacer

$M \leftarrow C$

$\text{CostoA} \leftarrow \text{CostoC}$

Intercambiar  $(m_1, m_i)$  si  $i \neq j$

*El centroide que se perturbo quede en la posición 1*

$i \leftarrow 2$

$j \leftarrow 0$

Fin Si

$j \leftarrow j+1$  Se va a perturbar con el siguiente objeto

Si  $j < n-k$  entonces

*Si ya pasaron a ser centroides todos los objetos en la posición i*

$i \leftarrow i+1$

si  $i = k+1$  entonces s

*i ← 1 si ya se perturbaron todos los objetos en todas las posiciones*

fin si

$j \leftarrow 1$

fin Si

fin para

$t \leftarrow \alpha t$

fin Mientras

Este algoritmo fue implementado en un lenguaje de alto nivel en una máquina con Procesador Intel Centrino® a 1.4 Ghz, Memoria de 768MB y Disco Duro de 80GB.

### Sección 3: Pruebas

La dosis que hemos impuesto a los parámetros es de carácter puramente experimental. A continuación se muestra la descripción de estos:

Temperatura inicial ( $T_i$ ): 5000. La  $T_i$  debe ser un valor alto para que la exploración en el espacio de soluciones sea aceptable. Alfa ( $\alpha$ ): 0.98. En cada iteración se multiplica la temperatura actual por alfa. Con este

mecanismo de enfriamiento, al principio (cuando la temperatura es muy alta) el sistema se enfría rápidamente (cuando hay mayor probabilidad de aceptar una solución peor que la actual), cuando la temperatura es baja, el valor de la temperatura actual decrece cada vez menos. Si el valor de alfa es muy pequeño, el sistema se enfriará rápidamente; entre mas cercano este a cero, el numero de exploraciones será mas pequeño aunque la temperatura inicial sea muy alta. Por ejemplo, con  $\alpha=0.1$  solo se realizarán alrededor de 4 iteraciones. Entre mas cercano este el valor de alfa a uno, el numero de iteraciones será mayor y por tanto se explorara mas espacio de soluciones. Temperatura final (Tf): se eligió a 0.001. En teoría el algoritmo debe finalizar cuanto la temperatura actual sea igual a cero, pero al ir enfriando el sistema multiplicando el valor de la temperatura actual con alfa, el valor de la temperatura actual se ira haciendo cada vez mas pequeño sin llegar a ser cero, por lo que un valor muy pequeño es adecuado como criterio de finalización.

La siguiente tabla revela resultados de solo algunas pruebas piloto. Como puede observarse, el agrupar las variables censales (datos poblacionales, Pob), hace que el número de soluciones aceptadas sea menor independientemente del método de recocido que se haya aplicado debido a son 57 variables, y con Coord. se agrupan los valores de coordenadas geográficas (minimización de distancias).

Dos algoritmos de Recocido simulado se probaron, el primero como se describe en la sección 2 (Recocido Simulado Aleatorio), el segundo con la variante de no generar soluciones que ya habían sido encontradas (Recocido Simulado Sistemático) [ ].

	NC	Ti	Tf	alpha	L(t)	TD	Ni	NSA
RSS	12	5000	0.001	0.98	5	Pob	3820	6
RSS	36	5000	0.001	0.98	5	Pob	3820	12
RSS	12	5000	0.001	0.99	9	Pob	13815	4
RSA	12	5000	0.1	0.98	5	Coord	2680	1251
RSA	36	5000	0.1	0.98	5	Coord	2680	1460

Tabla 1: Concentrado de pruebas.

Se eligieron dos pruebas de las obtenidas (2 últimos renglones señalados de la tabla anterior), para que al graficarse el número de iteraciones contra el costo de la función objetivo, se evidenciara la convergencia del algoritmo (Fig.1 y Fig.2). Se eligieron dos pruebas de las obtenidas en las instancias de la tabla anterior para que al graficarse el número de iteraciones contra el costo de la función objetivo, se observará la convergencia del algoritmo que hemos implementado. Las pruebas corresponden a 12 y 36 grupos respectivamente (Fig.1 y Fig.2).

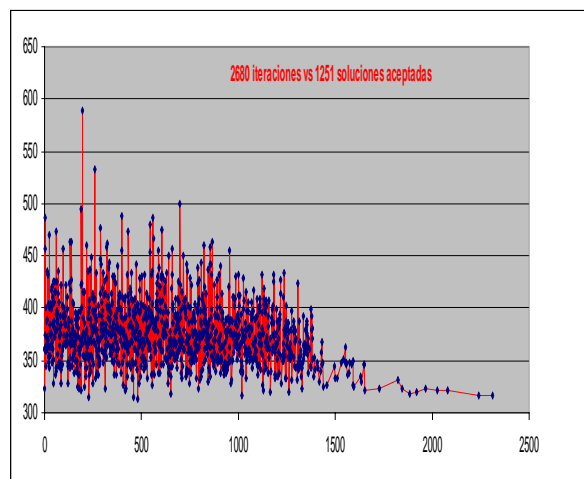


Figura 1: 12 grupos.  $T_i=5000$ ,  $T_f=.1$ ,  $\alpha=.98$

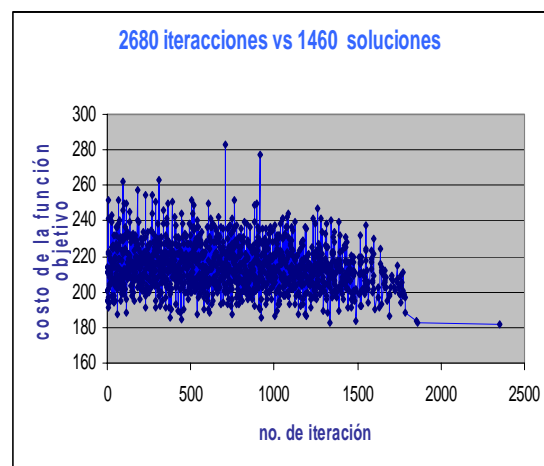


Figura 2: 36 grupos.  $T_i=5000$ ,  $T_f=.1$ ,  $\alpha=.98$



#### Sección 4: Conclusiones

Este trabajo ha permitido determinar que Recocido Simulado es una buena heurística para el problema de conglomerados sobre áreas geográficas.

Un desafío para los autores es insistir en encontrar un conjunto “elite” de soluciones que expresen como la agrupación de los agebs tiene un alto grado de compacidad geométrica cumpliendo con ciertas características poblacionales en un sentido de homogeneidad.

Para que los resultados puedan verse gráficamente, nos encontramos en un punto de concluir los detalles de la integración de un Sistema de Información Geográfica SIG a nuestro método de clasificación.

Finalmente estamos trabajando para encontrar la idoneidad de los parámetros de recocido simulado que permita determinar que tan cerca estamos de una buena solución.

#### Referencias

- [1]. Bernábe L., A. Osorio, J.C Duque “Clasificación Sobre Zonas Geográficas: Un Enfoque de Optimización Combinatoria para el Problema de Regionalización”. XIII CLAIO Congreso Latino-Iberoamericano de Investigación Operativa, Montevideo Uruguay, noviembre 2006.
- [2] Bernábe L.M.B, Aguirre V.R, & López S. R.; “Utilidad del Componente Principal para Caracterizar Rasgos Poblacionales” 2do.Congreso Latinoamericano de Matemáticos (UMALCA), Cancun, México, junio 2004.
- [3]. Cliff, A. D., Haggett, P., Ord, J. K., Bassett, K. A., and Davies, R. B. (1975). Elements of spatial structure : a quantitative approach. Cambridge [Eng.] ; New York : Cambridge University Press.
- [4]. Duque, J. C. (2004). Design of homogeneous territorial units. A Methodological Proposal and Applications. PhD dissertation, University of Barcelona.
- [5]. Fernando Bação, Victor Lobo, Marco Painho; “Applying genetic algorithms to zone design”; Springer Verlag; September 2004.
- [6]. Gordon AD (1999) Classification, 2nd ed. Boca Raton: Chapman & Hall-CRC
- [7]. J. Kalesics, S. Nickel, M. Schröder; “Towards a Unified Territory Design Approach – Applications, Algorithms and GIS Integration”, Universität des Saarlandes, Germany; January 2005.
- [8]. Kaufman, L. and P.J. Rousseeuw, Clustering by means of medoids, in: Y. Dodge (Ed.), Statistical Data Analysis based on the L1 Norm (North-Holland, Amsterdam, 1987) 405-416.
- [9]. Leberingber; Adaptive Simulated Annealing (ASA): lessons learned; Technical Report, Control and Cybernetics McLean VA 22101, 1995
- [10]. Loranca, M.B. & López, S., R. Aguirre V; “Application of Non-Supervised Classification to Population Data” ICEEE/CIE2004, International Conference on Electrical and Electronics Engineering, Acapulco Mexico, Septiembre 2004. ISBN 0-7803-8531-4.
- [11]. Mehrotra A, Johnson EL, Nemhauser GL (1998); “An optimization based heuristic for political districting”. Manage Sci. 44, pp. 1100–1114
- [12]. Murtagh F (1985) A survey of algorithms for contiguity-constrained clustering and related problems. Computer Journal 28:82-88
- [13]. Romero D., Burguete J., Martínez E., Velasco J. “Parcelación del territorio nacional: Un enfoque de optimización combinatoria para la construcción de marcos de muestreo en hogares”. INEGI, Agosto 2004.
- [14]. Rousseeuw, P.J., Hubert M., Struyf A. Clustering in an object-oriented environment. Department of Mathematics and Computer Science, U.I.A., Universiteitsplein 1, B-2610 Antwerp, Belgium.
- [15]. Zamora Alcocer Enrique “Implementación de un algoritmo compacto y homogéneo para la clasificación de zonas geográficas AGEBS bajo una interfaz gráfica”.
- [16] Instituto Nacional de Estadística, Geografía e Informática (INEGI). <http://www.inegi.gob.mx>
- [17] “Zonificación Óptima” B. Bernabe., Proyecto de tesis en desarrollo, Posgrado de Ingeniería UNAM, Investigación de Operaciones.

**Dirección del autor:** Maria Beatriz Bernábe Loranca  
email: [beatriz.bernabe@gmail.com](mailto:beatriz.bernabe@gmail.com)