

Propuesta y ajuste de un modelo de migración humana entre regiones colombianas empleando optimización

Proposal and adjusting a human migration model using optimization between Colombian regions

Lizeth Paola Martínez Ochoa*
Helbert Eduardo Espitia Cuchango**

Fecha de recepción: 27 de marzo de 2013

Fecha de aprobación: 30 de abril de 2013

Resumen

En este documento se propone un modelo dinámico para la migración de personas entre regiones colombianas. El modelo se establece considerando diagramas causales, posteriormente este modelo se ajusta empleando un algoritmo genético y un algoritmo basado en el cálculo de gradientes. Primero se emplea el algoritmo genético ya que este permite tener una buena exploración del espacio de búsqueda posteriormente con el algoritmo basado en gradiente se logra una buena convergencia. Los resultados muestran un buen ajuste del modelo.

Palabras clave

Modelo dinámico, identificación de parámetros, optimización, algoritmos genéticos.

* Universidad Distrital Francisco José de Caldas Proyecto Curricular de Ingeniería de Sistemas. Correo electrónico: heespitiac@udistrital.edu.co

** Universidad Distrital Francisco José de Caldas Proyecto Curricular de Ingeniería de Sistemas. Correo electrónico: lpmartinezo@correo.udistrital.edu.co

Abstract

This paper proposes a dynamic model for people migration between Colombian regions. The model is established considering causal diagrams, then fits this model using a genetic algorithm and an algorithm based on gradients. First genetic algorithm is used, this allows a good exploration of the search space and then with a algorithm based on gradients a good convergence is achieved.

Keywords

Dynamic model, parameter identification, optimization, genetic algorithms.

1. Introducción

La migración humana consiste en el movimiento o desplazamiento de personas desde un lugar de origen hacia un lugar de llegada, atravesando los límites de una división geográfica [1]. La migración de una determinada zona geográfica se compone de dos tipos de desplazamientos: la inmigración (llegadas) y la emigración (salidas).

Existen varios aportes teóricos para la concepción de una teoría de migración, entre ellos se encuentra las Leyes Migratorias de Ravenstein, la cuales fueron formuladas por Ernest George Ravenstein en el siglo XIX, quien consideró que la mayoría de las migraciones eran voluntarias; por lo tanto, los incentivos para migrar básicamente eran mejoras económicas, educacionales o climáticas [2].

2. Migración humana

Sobre los diferentes escenarios de investigación de la migración humana los más destacados son:

- Migración con fines económicos.
- Migración por aspectos sociales.

- Análisis migratorio haciendo uso de redes de información.
- Migración familiar.

2.1 Migración con fines económicos

Sobre estudios de migración humana entre zonas geográficas, en [3], [4] y [5] se pretende medir la población total de diferentes clases de personas migrantes (obrereros, estudiantes, etc.) tanto en la zona geográfica de origen como en la de destino en este trabajo se tiene en cuenta la utilidad del lugar de destino y el costo de migrar.

Por otro lado en [6], [7], [8] y [9] se realiza un estudio donde a partir de un modelo espacial se evalúa el efecto que tiene el valor de las casas o viviendas de cierta localidad en el cambio migratorio de la población, específicamente se estudia la población de trabajadores de oficina. Se tiene como resultado que un alza en el valor de las casas disminuye el flujo de inmigrantes en esta localidad, pero aumenta considerablemente en las localidades vecinas.

Finalmente, en [10] se investigan los factores determinantes en el crecimiento de los ingresos promedio y la tasa de migración

neta en 271 municipios de Suecia durante el periodo de 1981 y 1999.

2.2 Migración por aspectos sociales

Con respecto a la migración de personas por aspectos sociales, en [11] se proponen tres modelos estocásticos para estimar proyecciones poblacionales a nivel de mortalidad, fertilidad y migración neta internacional.

Sobre el caso colombiano, en [12] se mide la fuerza de expulsión o flujo de emigrantes de ciertas zonas de Colombia durante tres periodos determinantes correspondientes a los años 2000-2002, 2003-2005 y 2006-2008, en los cuales hubo transformaciones importantes en la actividad de los grupos armados.

Por otro lado en [13] se simula el movimiento de personas entre diferentes zonas; se emplea la teoría de partículas encontrando como resultado que esta técnica describe acertadamente el movimiento y el comportamiento de los individuos y grupos migrantes según las condiciones topográficas.

2.3 Análisis migratorio haciendo uso de redes de información

Como primer trabajo a considerar en [14] se utilizan los datos de mexicanos que migraron hacia U.S.A. durante el periodo de 1978 - 1998 para analizar empíricamente el papel que desempeñan las redes de información en la elección de una zona de destino para migrar.

En [15] se muestra la importancia de la información proporcionada por las redes familiares y comunales sobre los rendimientos esperados y los costos de migrar en la decisión de realizar futuras migraciones de México a Estados Unidos.

Por otro lado en [16] se estudia el efecto de las redes sociales en la disminución de los costos de migrar (y por ende un aumento en los beneficios de migrar) y la relación empí-

rica y teórica entre las emigraciones y la desigualdad rural en México.

2.4 Migración familiar

Un modelo dinámico sobre la decisión de migrar de parejas casadas se puede apreciar en [17], donde se utilizan los datos dados por el Panel de Estudios de Dinámicas de Ingreso (PSID) con el fin de obtener estimaciones sobre la relación entre el mercado laboral de los matrimonios y la migración.

Otro trabajo relacionado se aprecia en [18] en el cual se analiza la decisión de migrar en un contexto familiar, su influencia en próximas migraciones y por ende en los cambios de empleo y salario de los miembros de la familia.

Finalmente en [19] se analiza empíricamente el papel que desempeña la familia en las inmigraciones a Estados Unidos, teniendo en cuenta la comparación de las ganancias de cada miembro de la familia y el costo de migrar.

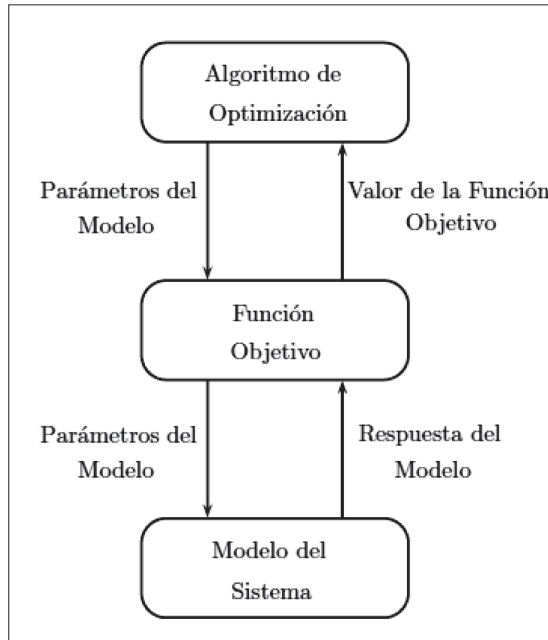
3. Propuesta para identificación de parámetros mediante optimización

El método empleado para realizar la identificación de parámetros del modelo se puede apreciar en la figura 1. Como es de apreciar en la figura primero se implementa una función en la que se calcula la respuesta dinámica del sistema con la cual se calcula el respectivo índice de desempeño el cual se optimiza.

3.1 Optimización por algoritmos genéticos

Los algoritmos genéticos GA se basan en el mecanismo empleado por la naturaleza para que una especie pueda adaptarse a un entorno [20]; estos algoritmos simulan los procesos de la evolución darwiniana mediante el uso de operadores genéticos que operan sobre una población, de individuos o cro-

Figura 1. Proceso para la identificación de parámetros del modelo



Fuente: elaboración propia.

mosomas, que evoluciona de una generación a otra.

Los algoritmos genéticos son muy empleados en problemas complejos de optimización con espacios de búsqueda muy grandes, donde pueden existir varios máximos o mínimos locales [20]. En los algoritmos genéticos se tiene dos elementos principales, el genotipo que consiste en la codificación empleada para cada individuo y el fenotipo el cual corresponde a la forma como el individuo se desempeña en un determinado entorno. Por lo general, los algoritmos genéticos emplean una codificación binaria para el genotipo y el fenotipo está dado por la función objetivo.

Los pasos en un algoritmo genético básico son:

1. Generar la población inicial de forma aleatoria.
2. Evaluar el desempeño de cada individuo.

3. Seleccionar los mejores individuos de forma probabilística.
4. Aplicar operador de cruce de forma probabilística.
5. Aplicar operador de mutación de forma probabilística.
6. Repetir desde el paso 2 hasta cumplir algún criterio de finalización.

En un algoritmo genético simple se tienen los operadores de cruce y mutación, los cuales, por lo general, se aplican de forma estocástica. El operador de cruce ayuda a tener una convergencia del algoritmo mientras que el operador de mutación incrementa la exploración del espacio de búsqueda.

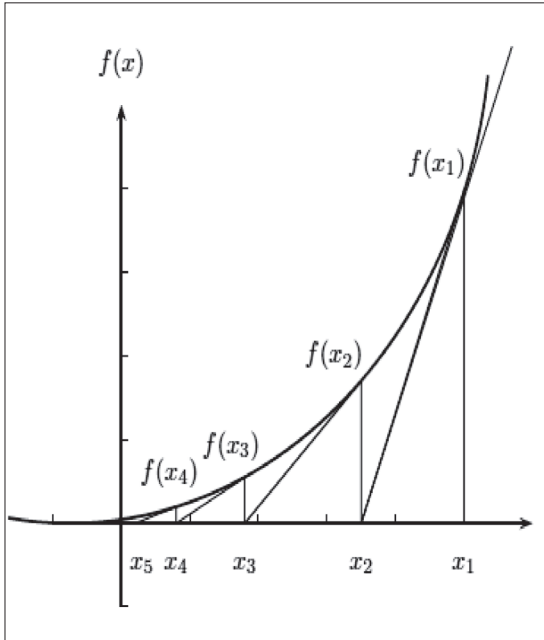
El cruce consiste en tomar dos individuos y combinar sus codificaciones para generar nuevos individuos, mientras que en el proceso de mutación se cambia uno o varios elementos de la codificación de un individuo [20].

3.2 Optimización sin restricciones

En un algoritmo de optimización sin restricciones para determinar el punto siguiente en una iteración se tienen dos estrategias: búsqueda de línea y región de confianza. En la estrategia de búsqueda de línea se escoge una dirección sobre la cual se busca un punto que proporcione un menor valor de la función objetivo, en cada iteración se busca una nueva dirección y un nuevo punto. En la estrategia de región de confianza se construye una función modelo que se aproxime a la función objetivo; dado que el modelo para ciertos puntos difiere de la función objetivo, se establece una región donde esta aproximación sea buena, la cual se denomina región de confianza [21].

Un ejemplo de un algoritmo con estrategia de búsqueda de línea es el método de cuasi-Newton donde se realizan aproximaciones sucesivas del Hessiano. Como ejemplos de

Figura 2. Proceso recursivo para encontrar la raíz de $f(x)$



Fuente: elaboración propia.

métodos basados en región de confianza se tienen los que emplean formas cuadráticas [21].

3.3 Método de Newton

El proceso iterativo para encontrar los ceros de una función empleando el método de Newton-Raphson se puede apreciar en la figura 2.

La ecuación recursiva de este método es:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \quad (1)$$

Para procesos de optimización se requiere encontrar las raíces de $f'(x)$, en tal caso la ecuación recursiva tiene la siguiente forma:

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad (2)$$

donde x_{k+1} es un punto que busca minimizar a f .

Cuando la función a minimizar es de varias variables el método de Newton toma la forma:

$$X_{k+1} = X_k - (f''(X_k))^{-1} f'(X_k) \quad (3)$$

Donde:

- f : función a minimizar.
- X : es el vector formado por las variables de f .
- f' : es el vector de primera derivadas de f también denominado gradiente.
- f'' : es la matriz hessiana cuyos elementos son las segundas derivadas de f .

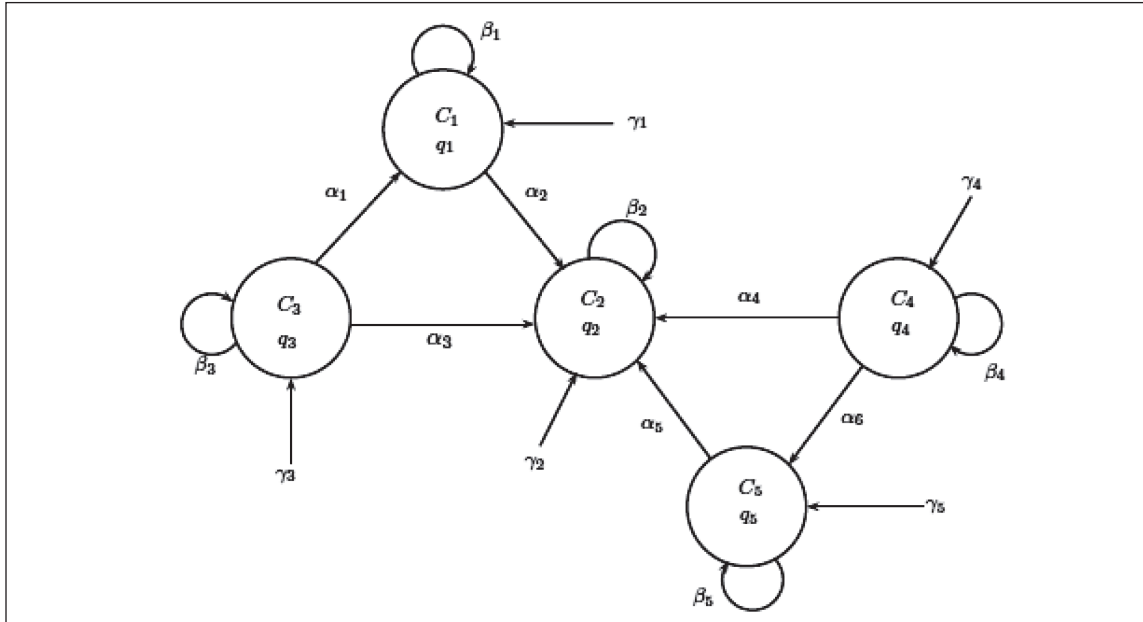
4. Modelo propuesto

El esquema del modelo propuesto para la migración entre cinco regiones colombianas (andina, caribe, orinoquia, amazonia y pacífica), representado con un diagrama causal se presenta en la figura 3.

Cabe destacar que las relaciones entre un par de regiones se suponen existentes, siempre y cuando estas regiones posean fronteras adyacentes. Considerando el diagrama mostrado en la figura 3 se obtiene el siguiente conjunto de ecuaciones:

$$\begin{aligned} C_1 \frac{dq_1}{dt} &= \beta_1 q_1 + \alpha_1 (q_3 - q_1) - \alpha_2 (q_1 - q_2) + \gamma_1 \\ C_2 \frac{dq_2}{dt} &= \beta_2 q_2 + \alpha_2 (q_1 - q_2) + \alpha_3 (q_3 - q_2) + \\ &\quad \alpha_4 (q_4 - q_2) + \alpha_5 (q_5 - q_2) + \gamma_2 \\ C_3 \frac{dq_3}{dt} &= \beta_3 q_3 - \alpha_1 (q_3 - q_1) - \alpha_3 (q_3 - q_2) + \gamma_3 \\ C_4 \frac{dq_4}{dt} &= \beta_4 q_4 - \alpha_4 (q_4 - q_2) - \alpha_6 (q_4 - q_5) + \gamma_4 \\ C_5 \frac{dq_5}{dt} &= \beta_5 q_5 - \alpha_5 (q_5 - q_2) + \alpha_6 (q_4 - q_5) + \gamma_5 \end{aligned} \quad (4)$$

Figura 3 . Diagrama causal propuesto para modelar los flujos migratorios en cinco regiones colombianas



Fuente: elaboración propia.

Con $i = 1, 2, 3, 4, 5$, donde:

- q_i corresponde a la región caribe, q_2 a la región andina, q_3 a la región pacífica, q_4 a la región orinoquia y q_5 a la región amazonia.
- C_i representa la capacidad que tiene la i -ésima región para almacenar personas.
- α_i pondera el flujo neto de personas entre dos determinadas regiones.
- γ_i es un flujo constante de personal, proveniente de otros lugares no especificados en el modelo al lugar i .
- β_i representa el crecimiento propio de la población en cada región y la migración entre zonas de menor tamaño (departamentos, municipios, etc.) pertenecientes a la misma región i .

Con el fin de tener un menor número de parámetros el anterior modelo se puede escribir como:

$$\begin{aligned}
 \frac{dq_1}{dt} &= a_{11}q_1 + a_{12}q_2 + a_{13}q_3 + b_1 \\
 \frac{dq_2}{dt} &= a_{21}q_1 + a_{22}q_2 + a_{23}q_3 + a_{24}q_4 + a_{25}q_5 + b_2 \\
 \frac{dq_3}{dt} &= a_{31}q_1 + a_{32}q_2 + a_{33}q_3 + b_3 \\
 \frac{dq_4}{dt} &= a_{42}q_2 + a_{44}q_4 + a_{45}q_5 + b_4 \\
 \frac{dq_5}{dt} &= a_{52}q_2 + a_{54}q_4 + a_{55}q_5 + b_5
 \end{aligned}
 \tag{5}$$

Con esta representación a_{ij} es el factor de influencia de migración de la región j a la región i ; del mismo modo, b_i es un término de ajuste entre las personas que ingresan y salen de otros lugares, los cuales no se consideran en el modelo. El factor de influencia puede ser una función dependiente de otros factores, sin embargo en esta propuesta se considera constante.

5. Función objetivo

La función objetivo considerada corresponde al error cuadrático medio definido como:

$$J(X) = \sum_{i=1}^R \frac{1}{N} \left[\sum_{n=1}^N (r[n, i] - s[n, i, X])^2 \right] \quad (6)$$

Donde J es la función objetivo, X el conjunto de parámetros del modelo, n el índice de tiempo discreto, r los datos reales medidos, s los datos obtenidos de la simulación del modelo, N el número total de datos tomados, $i = 1, 2, 3, 4, 5$ el índice de cada región y R el número total de regiones.

6. Resultados

La implementación del proceso de identificación de parámetros se realizó en MATLAB[®] con las respectivas herramientas especializadas. Los resultados obtenidos consideran la respuesta del modelo propuesto para las regiones caribe, pacífica, andina, orinoquía y amazónica, simulando el comportamiento de la migración presentada desde 1985 hasta el año 2005 y proyectando el posible comportamiento que se presentará desde el 2005 hasta el 2020. Para el ajuste del modelo

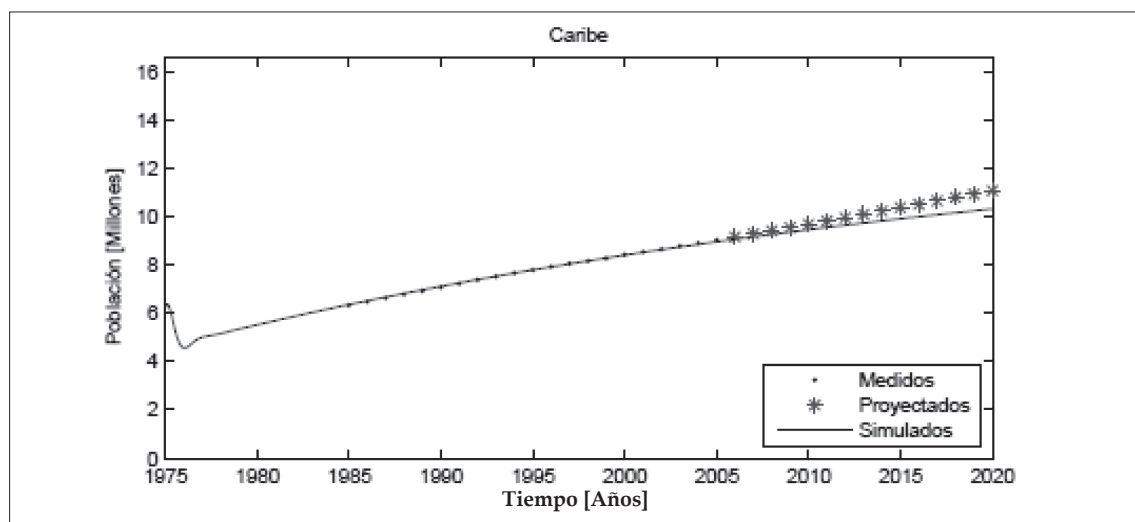
se utilizaron los datos suministrados por el DANE.

Como criterio de parada del algoritmo genético se tomó el número máximo de generaciones y para el algoritmo del gradiente se tomó el número máximo de evaluaciones de la función. Los parámetros del algoritmo genético se configuraron de la siguiente forma:

- Factor de Cruce 0.8.
- Generaciones: 1000.
- Rango inicial de los valores de la población: $[0, 1]$.
- Tamaño de la población: 20.
- Número de individuos que sobreviven a la siguiente iteración sin ningún cambio: 2.

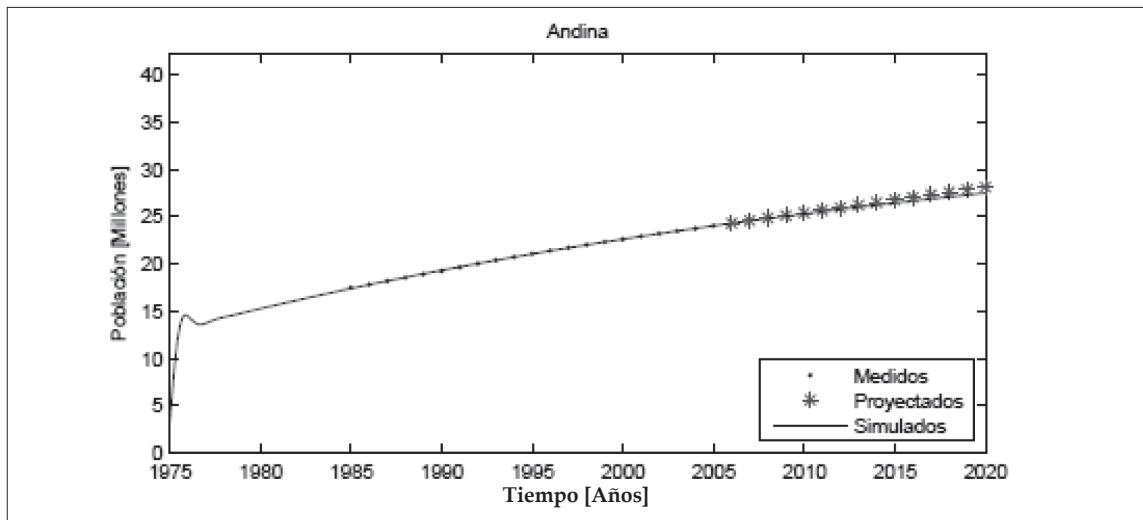
El proceso de optimización se corrió 50 veces, dado el comportamiento estocástico del algoritmo genético. Los valores de la función objetivo obtenidos en cada una de las iteraciones se presentan en la tabla 1. Es importante señalar que el resultado del algoritmo genético se emplea como punto de partida para el algoritmo de optimización basado en gradiente.

Figura 4. Resultado de simulación para la región Caribe



Fuente: elaboración propia.

Figura 5. Resultado de simulación para la región andina



Fuente: elaboración propia.

Como parámetros finales del modelo se eligen los que presentan un menor valor en la función objetivo con el algoritmo del gradiente. En las figuras 4, 5, 6, 7 y 8 se muestran los resultados del modelo para las regiones caribe, andina, pacífica, orinoquia y amazónica respectivamente. En estas figuras también se pueden apreciar los datos medidos y estimados por el DANE.

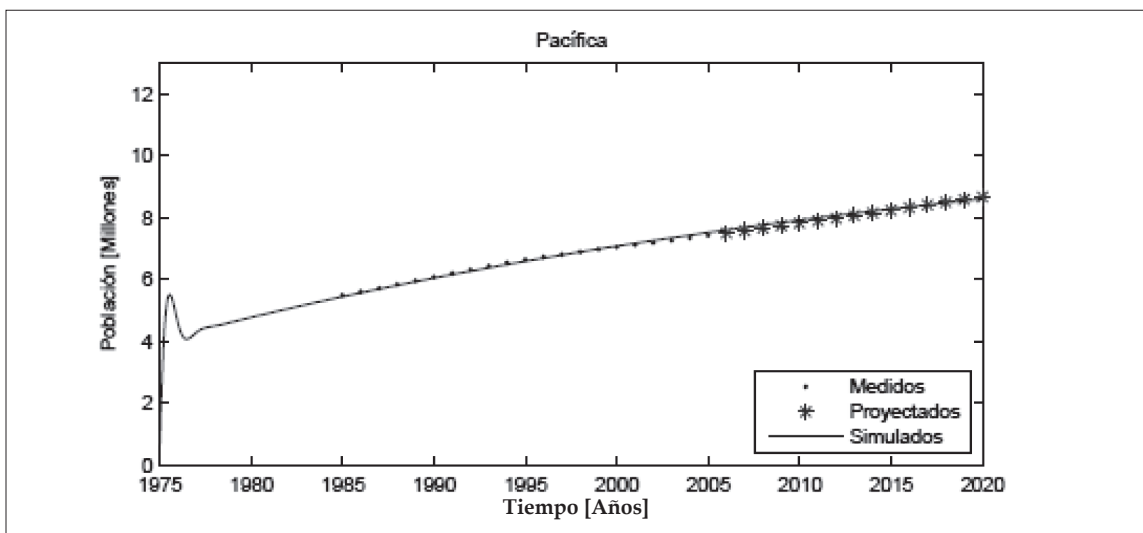
6.1 Modelo final

En el modelo obtenido se identificaron las siguientes condiciones.

$$y(0) = [6.19 \ 3.29 \ 0.69 \ 1.58 \ 2.11] \quad (7)$$

Considerando $dq/dt = q$ el modelo encontrado después del proceso de identificación de parámetros es:

Figura 6. Resultado de simulación para la región pacífica

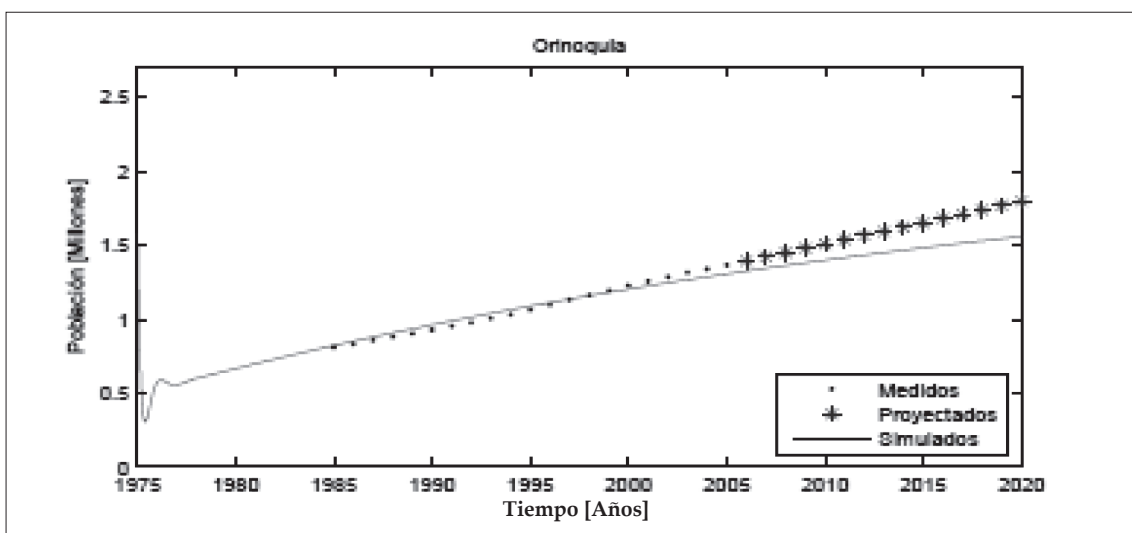


Fuente: elaboración propia.

Tabla 1. Resultados del proceso de optimización con la función

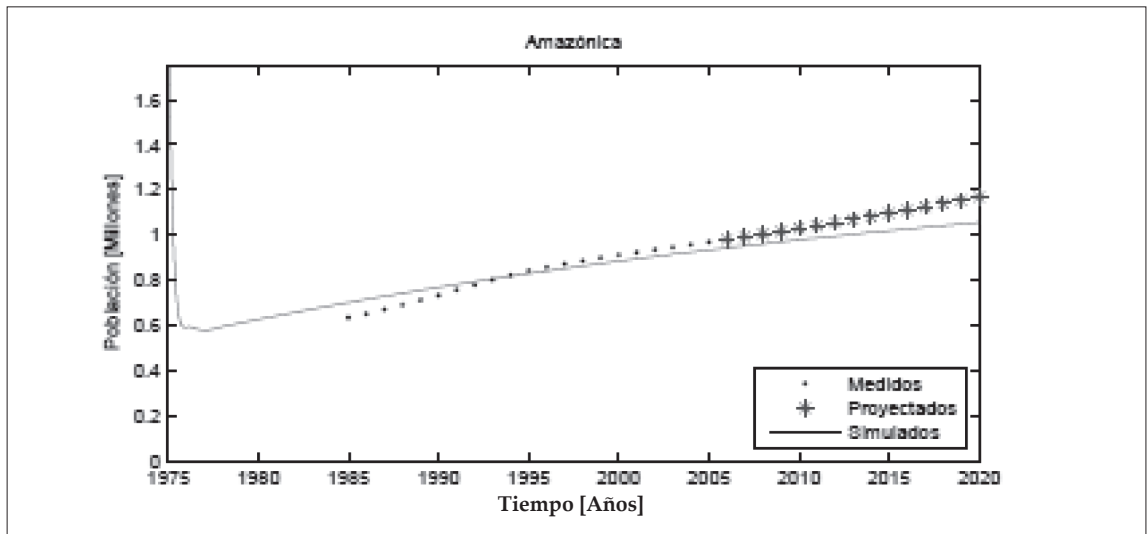
Ejecución	AG	Gradiente	Ejecución	AG	Gradiente
1	10,5794	0,5736	26	22,1177	0,4019
2	69,7232	0,8142	27	14,4450	3,5128
3	92,2175	0,4128	28	19,5327	5,9056
4	115,5289	113,7312	29	86,2544	84,9680
5	3,9824	0,3125	30	187,0816	0,0232
6	14,8316	4,6749	31	58,9348	0,0065
7	20,0316	0,0888	32	13,9924	3,8572
8	54,4415	4,8353	33	117,7461	2,7319
9	15,4696	5,0536	34	38,5189	3,5623
10	19,0307	4,5369	35	116,8952	0,0410
11	60,4419	60,4410	36	5,5132	0,2781
12	9,5284	0,0130	37	63,5791	0,2631
13	7,7305	4,6983	38	239,6169	2,2380
14	21,3462	0,1967	39	5,2881	0,0376
15	48,9992	1,7051	40	213,3657	0,3234
16	28,3844	0,0508	41	24,2966	0,2788
17	58,8877	0,1787	42	7,5294	5,2192
18	11,3765	0,2198	43	9,6695	3,6125
19	45,0144	0,0443	44	51,3548	0,0102
20	4,8480	0,0783	45	244,9127	3,4674
21	29,8487	0,0094	46	131,4903	73,1476
22	32,8522	5,1088	47	22,5196	0,5199
23	25,8939	5,0593	48	104,5179	0,1768
24	207,0273	1,1095	49	165,6724	0,1621
25	150,6448	0,0099	50	22,4967	0,0762

Fuente: elaboración propia.

Figura 7: Resultado de simulación para la región Orinoquía.


Fuente: elaboración propia.

Figura 8. Resultado de simulación para la región amazónica



Fuente: elaboración propia.

$$\begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \\ \dot{q}_3 \\ \dot{q}_4 \\ \dot{q}_5 \end{bmatrix} = \begin{bmatrix} 0.44 & 0.61 & -2.52 & 0 & 0 \\ 3.26 & -2.97 & 4.97 & 0.62 & 1.98 \\ 3.59 & -0.90 & -1.64 & 0 & 0 \\ 0 & 0.33 & 0 & -4.05 & -1.14 \\ 0 & 0.07 & 0 & 1.00 & -4.09 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \end{bmatrix} + \begin{bmatrix} 0.45 \\ 2.41 \\ 1.92 \\ -1.67 \\ 0.85 \end{bmatrix} \quad (8)$$

7. Conclusiones

A partir de los resultados cualitativos y cuantitativos, se puede apreciar que el modelo propuesto permite simular el comportamiento migratorio durante el tiempo medido (1985-2005), así como constatar los datos proyectados por el DANE (2005 - 2020) obteniendo como resultado un comportamiento análogo.

Para lograr un correcto ajuste del modelo fue necesario incluir las condiciones iniciales como variables de optimización y considerar un tiempo de holgura antes de realizar la comparación del resultado del modelo con los datos reales. Lo anterior se precisó, ya

que no es posible establecer una población ni un tiempo inicial para los años y datos reportados por el DANE.

Mediante el uso de diagramas causales fue posible proponer un modelo de migración para cinco regiones colombianas. Adicionalmente, a partir de este se logró determinar el conjunto de ecuaciones que describen el comportamiento migratorio presentado en el periodo estudiado.

La estrategia de optimización empleada, donde se mezclan algoritmos genéticos y minimización basada en gradiente, presenta un buen desempeño en la identificación de parámetros del algoritmo.

8. Referencias

- [1] Unión internacional para el estudio científico de la población, Diccionario demográfico multilingüe, Lieja, Bélgica, 1985.
- [2] G. Díaz, Aproximaciones metodológicas al estudio de las migraciones internacionales, Universidad Complutense de Madrid, UNISCI, Madrid, España, 2007.

- [3] L. Chavez, V. Kalashnikov, N. Kalashnycova, A human migration model with consistent conjectures, ACM Digital Library, México, 2009.
- [4] V. Kalashnikov, N. Kalashnycova, Simulation of a conjectural variation equilibrium in a human migration model, Congreso de Investigación y Desarrollo del Tecnológico de Monterrey, México, 2008.
- [5] V. Kalashnikov, N. Kalashnycova, A. Luévanos, R. Luévanos, M. Méndez, C. Uranga, Un modelo de migración humana: experimentos numéricos basados sobre los datos de las tres ciudades laguneras, Estudios Demográficos y Urbanos, Vol. 22, México, 2007.
- [6] P. Allen, Modeling the evolution of the U.S. spatial structure, Report No. DOT-CSSM-83-8, 1983.
- [7] W. Fang, L. Houqiang, Nonlinear Population Guideline, Sichuan University Publisher, 1995.
- [8] E. Irwin, W. Jeanty, M. Patridge, Estimation of a spatial simultaneous equation model of population migration and housing price dynamics, Regional Science and Urban Economics, Amsterdam, 2010.
- [9] J. Lu, D. Zengru, Self-organization Model of Population Space Distribution in PRC's towns, Academic paper of Beijing Normal University (science), Vol. 101-104, Beijing, 1990.
- [10] J. Lundberg, Using spatial econometrics to analyze local growth in Sweden, Centre for Regional Science (CERUM), University of Umea, Suecia, 2004.
- [11] H. Booth, R. Hyndman, Stochastic population forecast using functional data models for mortality, fertility and migration, Department of econometrics and Business Statistics, Monash University, ScienceDirect, Australia, 2008.
- [12] J. Sagayo, Desplazamiento forzoso en Colombia: expulsión y movilidad, dos dinámicas que interactúan, Universidad Nacional de Colombia, Facultad de Ciencias Económicas, Maestría en Ciencias Económicas, Colombia, 2011.
- [13] C. Christou, Simulation of Human Migration Based on Swarm Theory, Electronic Systems and technologies technical center, The MITRE Corporation, Estados Unidos de America, 2010.
- [14] L. Baghdadi, Mexico - U.S. Migration: Do spatial networks matter?, Université Paris I., mimeograph, 2005.
- [15] A. De Janvry, E. Sadoulet, P. Winters, Family and community networks in México - U.S. Migration, Agricultural and resource economics, University of New England, 1999.
- [16] D. Mckenzie, H. Rapoport, Network effects and the dynamics of migration and inequality: theory and evidence from Mexico, Journal of Development Economics, 2006.
- [17] A. Gemici, Family migration and labor market outcomes, University of Pennsylvania, Pennsylvania, 2006.
- [18] J. Mincer, Family migration decision, Center for Economic Analysis of Human Behavior and Social Institutions, New York, 1977.
- [19] G. Borjas, S. Bronars, Immigration and the family, The national Bureau of Economic research, 1990.
- [20] T. Weise, Global optimization algorithms - theory and application, Self-Published Thomas Weise, 2009.
- [21] H. Mora, Optimización no lineal y dinámica, Universidad Nacional De Colombia, 2001.