

Análisis de funciones de desempeño para el ajuste de parámetros de un modelo migratorio colombiano

Analysis of performance functions for
adjusting parameters of Colombian
migration model

Lizeth Paola Martínez*

Helbert Eduardo Espitia**

Universidad Distrital Francisco José de Caldas (Colombia)

* Ingeniera de Sistemas, Universidad Distrital Francisco José de Caldas, Proyecto Curricular de Ingeniería de Sistemas. paolizm@udistrital.edu.co

** Docente de planta, Universidad Distrital Francisco José de Caldas, Proyecto Curricular de Ingeniería de Sistemas. Doctor en Ingeniería de Sistemas y Computación. heespitiac@udistrital.edu.co

Correspondencia: Helbert Eduardo Espitia, Universidad Distrital Francisco José de Caldas, carrera 7 n°. 40B - 53 Piso 5 (Ing. Sistemas), Bogotá, Colombia.

Resumen

En este documento se propone el ajuste de parámetros para un modelo de migración humana entre cinco regiones colombianas. Para la identificación de parámetros se emplea un enfoque basado en optimización por lo cual se proponen dos funciones objetivo. Los resultados obtenidos con las funciones objetivo se analizan estadísticamente. La metodología para el análisis de datos consiste en emplear pruebas paramétricas o no paramétricas dependiendo del cumplimiento de los supuestos de normalidad y homocedasticidad (igualdad de varianza). Para el ajuste del modelo se emplean dos enfoques de la función objetivo, el primero considera el número total de la población y el segundo emplea la cantidad de habitantes de cada región. Realizando los análisis estadísticos se observa un mejor ajuste del modelo cuando se considera el número de habitantes de cada región.

Palabras clave: análisis estadístico, modelo de migración, optimización.

Abstract

This paper proposes the identification parameter for a human migration model among five Colombian regions. For the identification of parameters an optimization approach is employed where we propose two objective functions. The results obtained with the objective functions are statistically analyzed. The methodology for data analysis employs parametric or non-parametric tests depending on the fulfillment of the assumptions of normality and homoscedasticity (equal variance). To adjust the model, two approaches to the objective function are used. The first considers the total number of population and the second uses the number of inhabitants in each region. By carrying out statistical analyzes, a better adjustment model is observed when considering the number of inhabitants in each region.

Keywords: Model migration, optimization, statistical analysis.

Fecha de recepción: 17 de octubre de 2014
Fecha de aceptación: 10 de diciembre de 2015

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 se compone de dos tipos de desplazamientos: la inmigración, que es la llegada a una zona geográfica, y la emigración, que es la salida de una zona geográfica.

Existen varios aportes para la teoría de migración, entre ellos se encuentra las Leyes Migratorias de Ravenstein, las 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].

Ravenstein formuló una serie de leyes sobre la migración a partir del estudio del fenómeno en Inglaterra y Gales utilizando los datos obtenidos en el censo de 1881; posteriormente, en 1889 se amplió su investigación a 20 países más, comprobando las leyes antes formuladas y encontrando algunas nuevas [3].

Ravenstein determinó, a partir de los datos que analizó, que condiciones similares producían movimientos migratorios similares con características comunes entre las cuales se destacaban:

- Las migraciones se dan principalmente por condiciones económicas.
- Las migraciones se realizan por etapas.
- La mayoría de personas migrantes son adultos nativos de áreas rurales.
- Las grandes ciudades crecen más por personas que entran a la ciudad, provenientes de otras ciudades que por nacimientos.
- Las migraciones tienden a aumentar conforme al desarrollo económico y tecnológico.
- Los hombres son los principales migrantes.

Un modelo de migración humana permite establecer estrategias adecuadas para satisfacer las necesidades básicas de las personas en una determinada región geográfica.

En este trabajo se busca identificar los parámetros de un modelo de migración humana entre cinco regiones colombianas. Para realizar el ajuste del modelo se utilizan algoritmos de optimización por lo cual se proponen dos funciones de desempeño para esto. Considerando que puede existir diferencia entre los modelos obtenidos con las funciones de ajuste empleadas, entonces, se realiza un análisis estadístico para determinar la función de ajuste más adecuada.

Para la identificación de los parámetros del modelo se emplean datos reales proporcionados por el Departamento Administrativo Nacional de Estadísticas DANE (censos realizados desde 1985 hasta 2005).

MIGRACIÓN HUMANA

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

- Migración con fines económicos.
- Migración por aspectos sociales.
- Migración familiar.
- Influencia de la migración en la inversión directa extranjera.

Migración con fines económicos

Sobre estudios de migración humana entre zonas geográficas, en [4] se pretende medir la población total de diferentes clases de personas migrantes (obreros, estudiantes, etc.) tanto en la zona geográfica de origen como en la de destino.

Por otro lado, en [5] se presenta un modelo matemático operativo que simula los flujos migratorios internos de México y genera escenarios de migración

hipotéticos de acuerdo al comportamiento de variables independientes como: costo de migración, flujos migratorios de origen y capacidad de atracción.

Un trabajo adicional se observa en [6], donde se presenta un modelo escalonado de migración entre tres tipos de regiones (pueblos pequeños, ciudades no capitales y capitales), se tienen en cuenta factores como los costos del transporte, la cantidad poblacional en cada región, el mercado interno comparado con el externo, los niveles de proteccionismo del mercado regional y los beneficios de migrar, los cuales son medidos usando la relación entre el salario de la región de origen y el de destino.

En [7] 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 población de trabajadores de oficina. Sobre otros trabajos relacionados, en [8] se propone un modelo donde se incluye la existencia de migrantes permanentes los cuales corresponden a personas que trabajan en sectores diferentes a su lugar de origen pero que retornan a él frecuentemente. Este tipo de migración ocurre por las oportunidades de empleo y salarios ofrecidos en zonas aledañas.

Migración por aspectos sociales

Con respecto a la migración de personas por aspectos sociales, en [9] se proponen tres modelos estocásticos para estimar proyecciones poblacionales a nivel de mortalidad, fertilidad y migración neta internacional.

Por otro lado, el trabajo realizado en [10] presenta la simulación del movimiento de personas entre diferentes zonas. Para esto se emplea la teoría de partículas teniendo como resultado una adecuada simulación del movimiento y comportamiento de los individuos como también los grupos de migrantes según las condiciones topográficas.

Adicionalmente, en [11] se analiza el efecto de las anomalías del clima en los flujos de migración rural y urbana del África Sub-Sahariana, para esto se plantea un modelo empírico, con el cual se realizan estimaciones sobre el número de migrantes en África entre los años 1960 y 2000, así como proyecciones sobre las futuras migraciones en zonas donde la agricultura es la actividad predominante.

Otro estudio de este tipo se puede encontrar en [12] donde se realiza un modelo dinámico para explicar la distribución espacial y temporal de dos grupos de residentes en un área urbana, asumiendo que el comportamiento de cada grupo está determinado por factores económicos y sociales.

Sobre el caso colombiano, en [13] se mide la fuerza de expulsión o flujo de emigrantes de ciertas zonas de Colombia durante tres períodos determinantes 2000-2002, 2003-2005 y 2006-2008, en los cuales hubo transformaciones importantes en la actividad de los grupos armados. El modelo estima la intensidad de llegada (flujo de inmigrantes) a las diferentes zonas del país, determinando los municipios que reciben mayor cantidad de personas durante los períodos estudiados, según los incentivos económicos, la cantidad de población y la distancia que presenta cada una de estas zonas.

Migración familiar

Un modelo dinámico sobre la decisión de migrar de parejas casadas se puede apreciar en [14], 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 [15], en el cual se analiza la decisión de migrar en un contexto familiar, su influencia en próximas migraciones como también en los cambios de empleo y salario de los miembros de la familia. Adicionalmente, en [16] se analiza empíricamente el papel que juega 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.

Influencia de la migración en la inversión extranjera

Sobre trabajos enfocados en establecer la influencia que tiene la migración en la economía, en [17] se realiza un estudio para determinar la distribución regional de la inversión extranjera directa. Particularmente se estudia la inmigración de diez países a Estados Unidos y la correlación positiva que existe entre estos dos aspectos durante los años 1990 y 2004.

Otro trabajo a considerar se puede observar en [18] donde se visualiza que la existencia de redes de trabajadores de inmigrantes con estudios de bachillerato en Estados Unidos afecta positivamente la inversión extranjera directa

de Estados Unidos hacia 56 países de origen. Para esto se emplean variables como la cantidad de migrantes, los costos de obtener un pasaporte en el país de origen y la densidad de población en el país de origen de los migrantes.

Finalmente, en [19] se investiga si los flujos migratorios de trabajadores y la inversión extranjera directa son complementarios o existe una relación de sustitución entre ellos, esto se realiza por medio de un modelo estilizado con el cual se capturan los diferentes mecanismos para relacionar el mercado laboral y el movimiento del capital.

METODOLOGÍA PARA IDENTIFICACIÓN DE PARÁMETROS MEDIANTE OPTIMIZACIÓN

El método empleado desde una perspectiva general para realizar la identificación de parámetros del modelo se puede observar en la Figura 1.

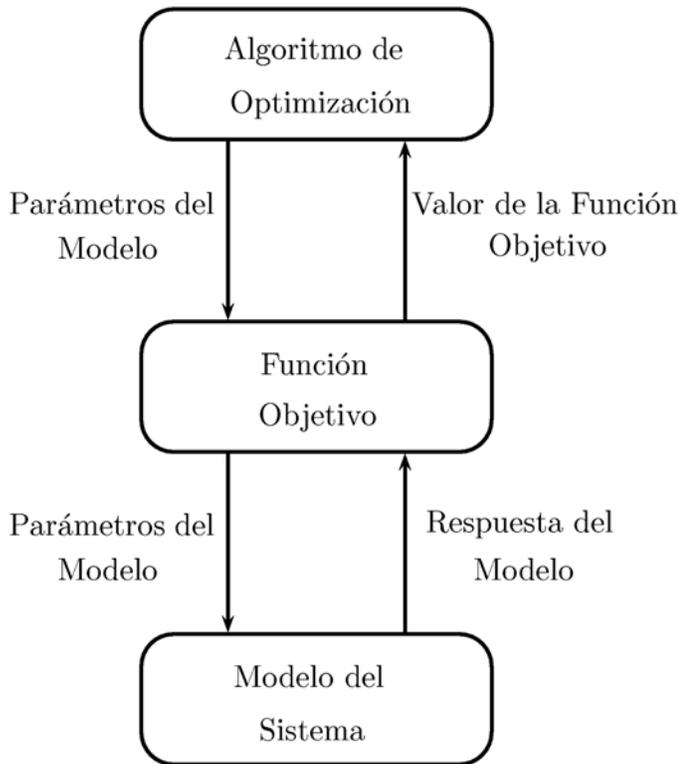


Figura 1. Esquema empleado para la identificación de parámetros del modelo.

Como es de apreciar en la Figura 1, en el nivel inferior se calcula la respuesta dinámica del sistema con la cual se establece el valor de la función objetivo siendo esta objeto de optimización. La estrategia de optimización propuesta emplea en una primera instancia algoritmos genéticos para tener una buena exploración del espacio de búsqueda y así encontrar una solución aproximada la cual se refina empleando el método de Newton.

MODELO PROPUESTO

Las regiones colombianas consideradas en el modelo son: Andina, Caribe, Orinoquia, Amazonia y Pacífica. En la Figura 2 se muestra el diagrama de relaciones empleado para representar la migración entre estas regiones.

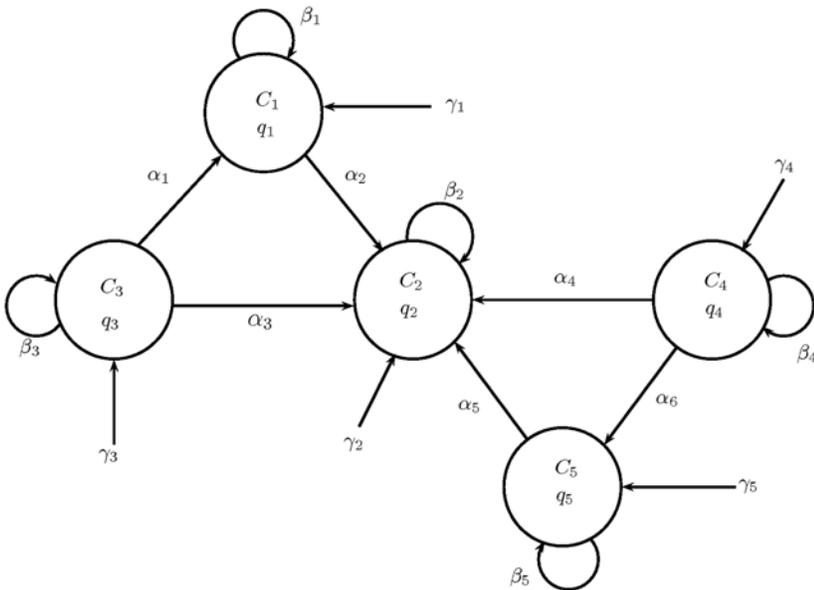


Figura 2. Diagrama de los flujos migratorios entre cinco regiones colombianas.

En este modelo se considera el flujo migratorio entre un par de regiones cuando estas presentan fronteras adyacentes. Tomando el diagrama mostrado en la Figura 2 se obtiene el siguiente conjunto de ecuaciones:

$$\begin{aligned}
 C_1 \frac{dq_1}{dt} &= \beta_1 q_1 + \alpha_1 (q_3 - q_1) - \alpha_2 (q_1 - q_2) + \gamma_1 \\
 C_2 \frac{dq_2}{dt} &= \beta_2 q_2 + \alpha_2 (q_1 - q_2) - \alpha_3 (q_3 - q_2) + \alpha_4 (q_4 - q_2) + \alpha_5 (q_5 - q_2) + \gamma_2 \\
 C_3 \frac{dq_3}{dt} &= \beta_3 q_3 + \alpha_1 (q_3 - q_1) - \alpha_3 (q_3 - q_2) + \gamma_3 \\
 C_4 \frac{dq_4}{dt} &= \beta_4 q_4 + \alpha_4 (q_4 - q_2) - \alpha_6 (q_4 - q_5) + \gamma_4 \\
 C_5 \frac{dq_5}{dt} &= \beta_5 q_5 + \alpha_5 (q_5 - q_2) - \alpha_6 (q_4 - q_5) + \gamma_5
 \end{aligned} \tag{1}$$

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

- q_1 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_1 pondera el flujo neto de personas entre dos determinadas regiones.
- γ_1 es un flujo constante de personal, proveniente de otros lugares no especificados en el modelo al lugar i .
- β_1 representa el crecimiento 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:

$$\begin{aligned}\frac{dq_1}{dt} &= \alpha_{11}q_1 + \alpha_{12}q_2 + \alpha_{13}q_3 + b_1 \\ \frac{dq_2}{dt} &= \alpha_{21}q_1 + \alpha_{22}q_2 + \alpha_{23}q_3 + \alpha_{24}q_4 + \alpha_{25}q_5 + b_2 \\ \frac{dq_3}{dt} &= \alpha_{31}q_1 + \alpha_{32}q_2 + \alpha_{33}q_3 + b_3 \\ \frac{dq_4}{dt} &= \alpha_{42}q_2 + \alpha_{44}q_4 + \alpha_{45}q_5 + b_4 \\ \frac{dq_5}{dt} &= \alpha_{55}q_5 + b_5\end{aligned}\tag{2}$$

En 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.

Para determinar la respuesta dinámica del sistema se requieren las condiciones iniciales de las poblaciones lo cual implica tener registros poblacionales desde un determinado punto histórico el cual se considera como el inicio del proceso de migración y crecimiento poblacional de las regiones, lo cual resulta difícil de establecer ya que en la práctica se tienen datos desde 1985. Por lo anterior, dado que se desconocen las condiciones iniciales del modelo estas se consideran como variables de optimización, de la misma forma al no precisar el inicio de la respuesta del sistema se toma un tiempo de holgura antes de realizar la comparación de los datos reportados por el Departamento Administrativo Nacional de Estadísticas (DANE).

Optimización mediante algoritmos genéticos

Los algoritmos genéticos (Genetic Algorithms, GA por sus siglas en inglés) se basan en el mecanismo empleado por la naturaleza para que una especie

pueda adaptarse a un entorno [20]. Los GA 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 las características del individuo en un determinado entorno. Por lo general, los algoritmos genéticos emplean una codificación binaria para el genotipo y el fenotipo corresponde a las características del individuo las cuales se evalúan en la función objetivo (entorno).

El fundamento de los algoritmos genéticos consiste en una población de individuos los cuales coexistente en un entorno con recursos limitados, de tal forma que la competencia por los recursos permite la selección de aquellos individuos que están mejor adaptados al entorno, los cuales se convierten en los padres de nuevos individuos mediante procesos de cruce y mutación. Con el paso del tiempo, este proceso de selección mejora el desempeño de los individuos de la población [21]. Los pasos básicos de un algoritmo genético se pueden apreciar en la Figura 3.

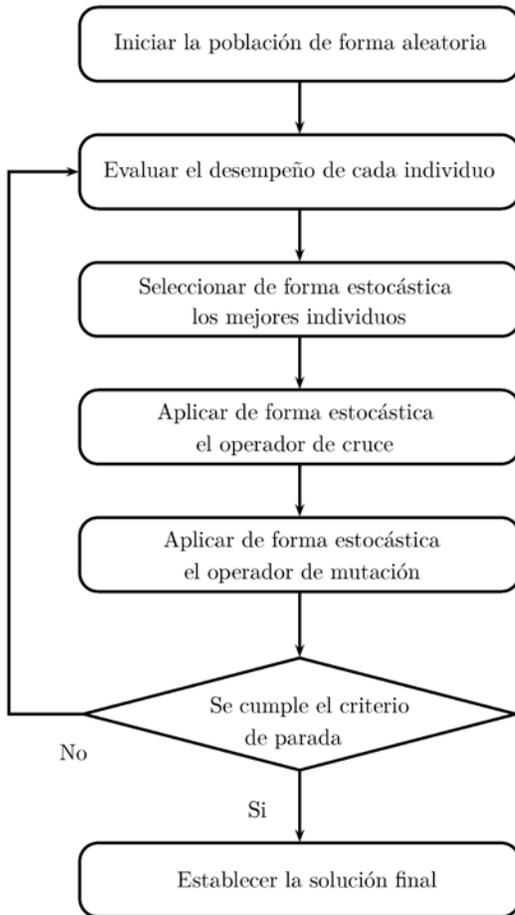


Figura 3. Proceso de un algoritmo genético [21].

Optimización sin restricciones

En un algoritmo de optimización sin restricciones se pueden tener dos estrategias para determinar el punto siguiente, estos enfoques son: búsqueda de línea y región de confianza. Para la primera, de forma iterativa se busca una dirección sobre la cual se tenga un punto que mejore la función objetivo. Para la segunda estrategia se construye una función modelo que se aproxime a la función objetivo de tal forma que se establece una región de confianza donde esta aproximación sea buena [22].

El método de cuasi-Newton es un ejemplo de un algoritmo con estrategia de búsqueda de línea. Por su parte, sobre algunos métodos basados en región de confianza se tienen los que emplean formas cuadráticas [22].

Método de Newton

En el método de Newton el proceso iterativo para determinar el valor mínimo para una función de una variable se encuentra sujeto a la siguiente ecuación:

$$x_{k+1} = x_k - \frac{f'(C)}{f''(x_k)} \quad (3)$$

Siendo x_{k+1} el punto que minimiza a f . Para una función de varias variables la ecuación asociada al método de Newton es:

$$X_{k+1} = X_k - (F''(X_k)^{-1} F'(X_k)) \quad (4)$$

Donde, X corresponde a un vector formado por las variables de decisión, F' es el vector de primeras derivadas (gradiente) y F'' es una matriz de segundas derivadas (matriz Hessiana).

Para la implementación del método de Newton el vector gradiente y la matriz Hessiana se calculan mediante aproximaciones numéricas de primeras y segundas derivadas respectivamente [22]. Como es de apreciar, también se requiere calcular la inversa de la matriz Hessiana lo cual puede ser costoso desde el punto de vista computacional, por lo cual existen métodos que realizan aproximaciones de esta inversa. Este conjunto de métodos se denominan cuasi-Newton de los cuales se tiene el método DFP (Davidon Fletcher Powell) y el método BFGS (Broyden Fletcher Goldfarb Shannon) [22].

FUNCIÓN OBJETIVO

Para una primera función objetivo el índice de desempeño considerado corresponde al error cuadrático medio definido a continuación:

$$J_1(X) = \sum_{i=1}^R \frac{1}{N} \left[\sum_{n=1}^N (r[n, 1] - s[n, i, X])^2 \right] \quad (5)$$

Dónde J_1 es la función objetivo, X el conjunto de parámetros del modelo, n la variable asociada al tiempo (discreto), r datos reales, s 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.

Considerando que las regiones no tienen la misma cantidad de personas, se propone una función objetivo adicional, en la cual se pondera el error para cada región tomando el número total de individuos de cada región. Esta función objetivo es:

$$J_2(X) = \sum_{i=1}^R \frac{1}{N} \left[\sum_{n=1}^N (r[n, 1] - s[n, i, X])^2 \right] / S(i) \quad (6)$$

Donde:

$$S(i) = \frac{1}{N} \sum_{m=1}^N r[m, 1] \quad (7)$$

Como es de apreciar, para la función objetivo J_2 se realiza la normalización considerando la población de cada región, por lo tanto, se espera tener un mejor ajuste con esta función objetivo lo cual se busca evidenciar con el análisis estadístico de resultados.

RESULTADOS

En esta sección se presentan los resultados para las dos funciones objetivo planteadas anteriormente. Los resultados obtenidos, para cada una de las funciones propuestas consideran la simulación del modelo 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, por otro lado, 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.
- Factor de Mutación 0.01.
- 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.

Para la selección de parámetros del AG se debe tener presente que el operador de cruce permite la convergencia a un valor óptimo mientras que el operador de mutación proporciona la capacidad de explorar otras regiones del espacio de búsqueda. Se debe tener presente que un valor bajo de cruce puede hacer que el algoritmo se demore en determinar un valor óptimo, por otro lado, un valor alto de mutación puede ocasionar problemas en la convergencia del algoritmo. Para el operador de cruce convencionalmente se emplean probabilidades cercanas al 90% mientras que para el operador de mutación se toman probabilidades menores del 1% [23].

Dado el comportamiento estocástico del algoritmo genético el proceso de optimización se corrió 50 veces. Es importante señalar que el resultado del algoritmo genético se emplea como punto de partida para el algoritmo de optimización basado en gradiente.

Resultados para J_1

Para esta función objetivo, el resumen de los valores obtenidos para las 50 ejecuciones se presenta en la Tabla 1.

Tabla 1. Resultados del proceso de optimización para J_1 .

Medida	GA	Newton
Máximo	244,9127	113,7312
Mínimo	3,9824	0,0065
Promedio	62,92	8,1756
Desviación estándar	66,6493	23,0913

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 Amazonia, respectivamente. En estas figuras también se pueden apreciar los datos medidos y estimados por el DANE.

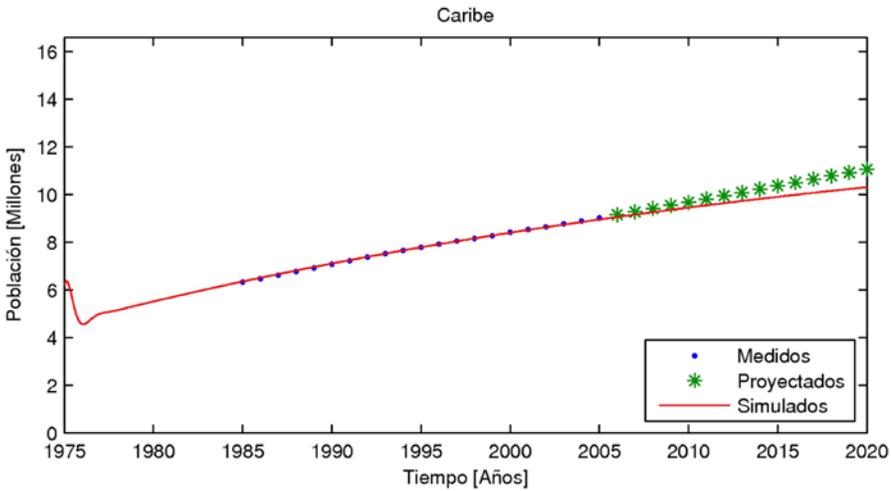


Figura 4. Resultado de simulación para la región Caribe con la función J_1 .

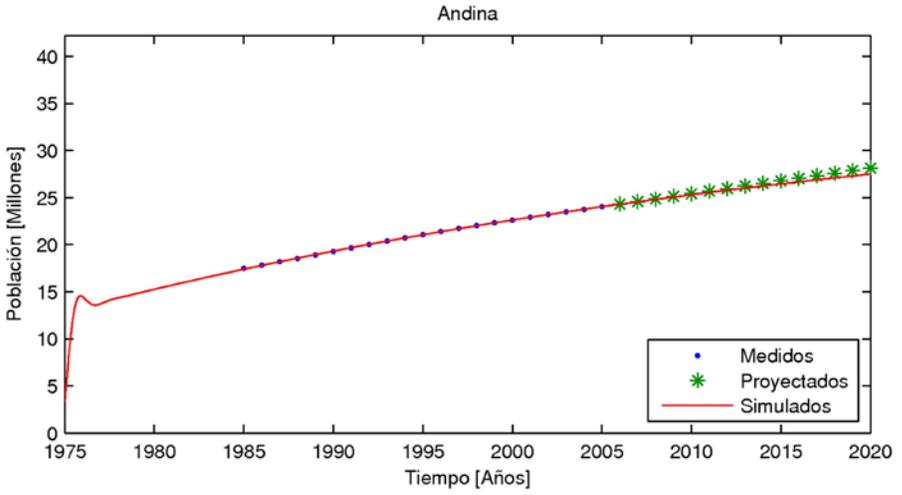


Figura 5. Resultado de simulación para la región Andina con la función J_1 .

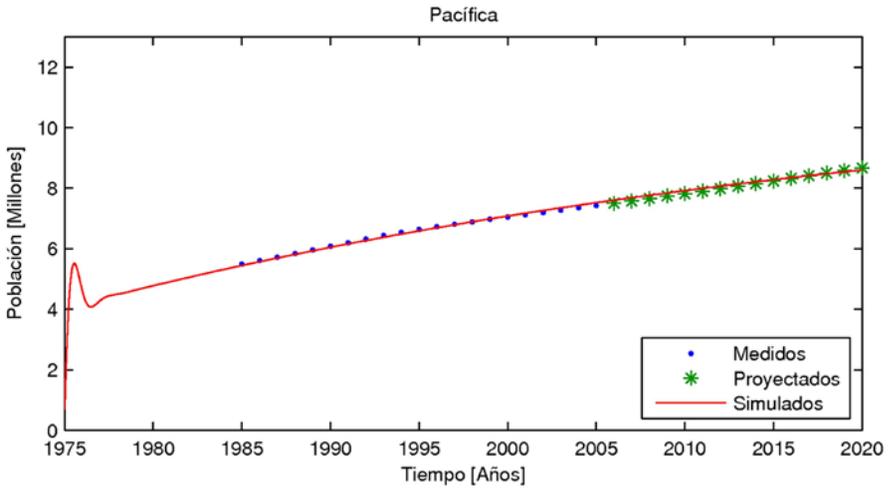


Figura 6. Resultado de simulación para la región Pacífica con la función J_1 .

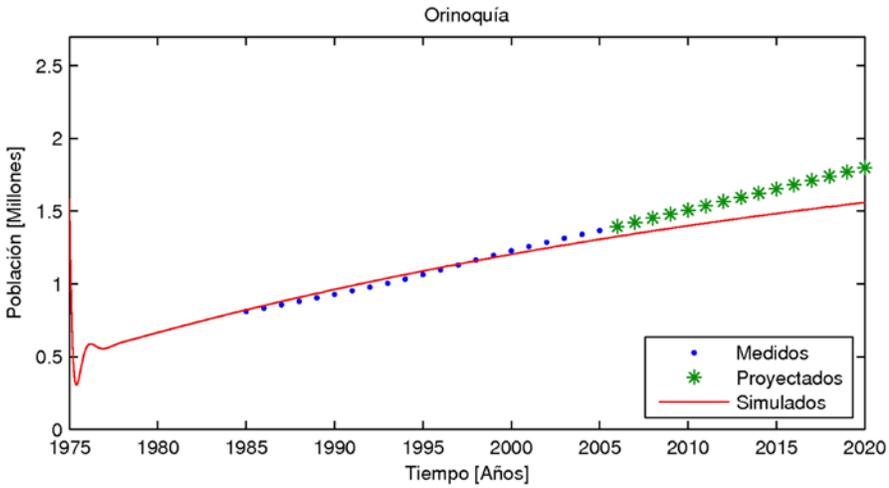


Figura 7. Resultado de simulación para la región Orinoquia con la función J_1 .

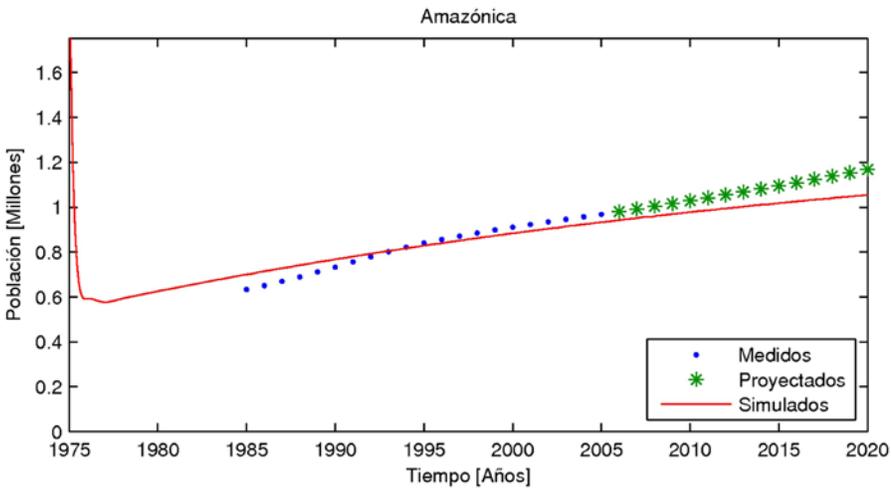


Figura 8. Resultado de simulación para la región Amazonia con la función J_1 .

Resultados para J_2

Al igual que en el caso anterior, para la función J_2 el proceso de optimización se corrió 50 veces. Como parámetros del modelo se toman los que presentan

un mejor valor de la función objetivo. El resumen de los valores obtenidos de la función objetivo para las 50 corridas se muestra en la Tabla 2.

Tabla 2. Resultados del proceso de optimización para J_2 .

Medida	GA	Newton
Máximo	23,4972	20,1606
Mínimo	0,3777	0,0026
Promedio	10,5748	1,441
Desviación estándar	7,4528	4,248

Las Figuras 9, 10, 11, 12 y 13 muestran los resultados obtenidos empleando la función objetivo J_2 para las regiones Caribe, Andina, Pacífica, Orinoquia y Amazonia, respectivamente.

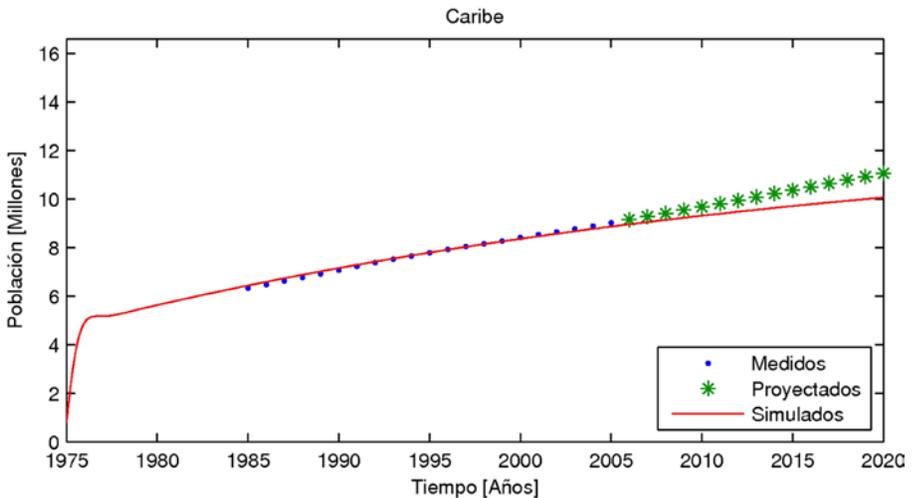


Figura 9. Resultado de simulación para la región Caribe con la función J_2 .

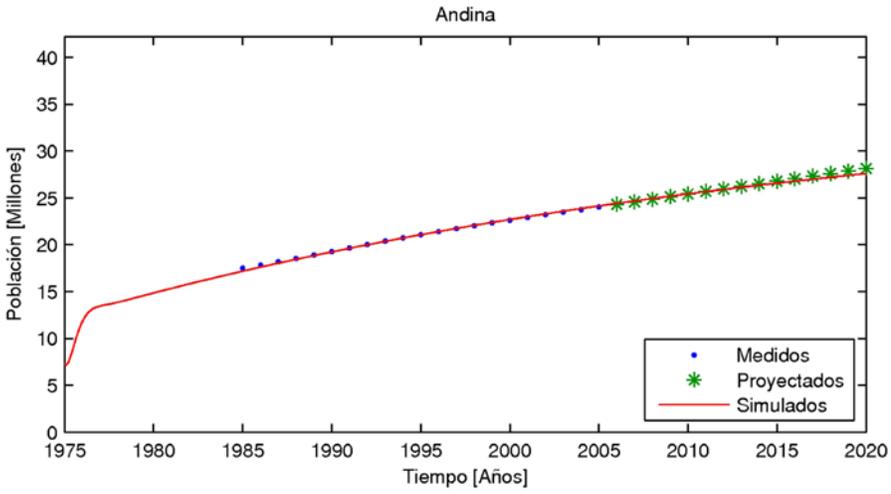


Figura 10. Resultado de simulación para la región Andina con la función J_2 .

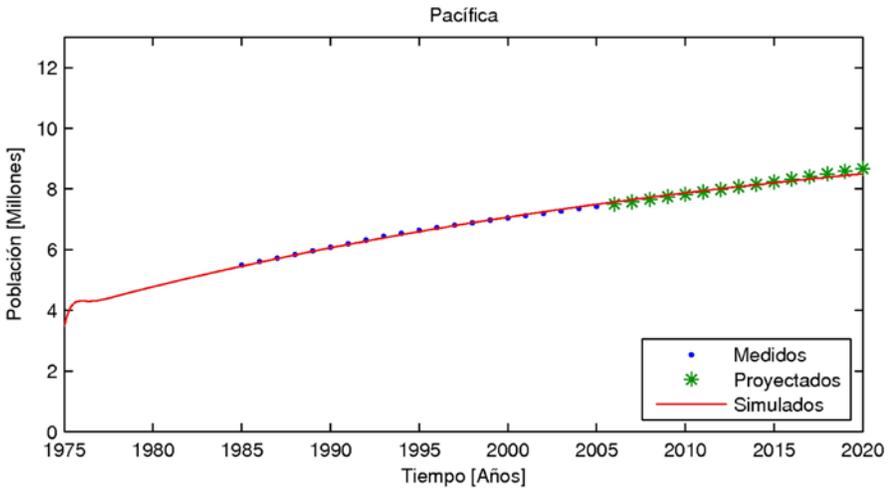


Figura 11. Resultado de simulación para la región Pacífica con la función J_2 .

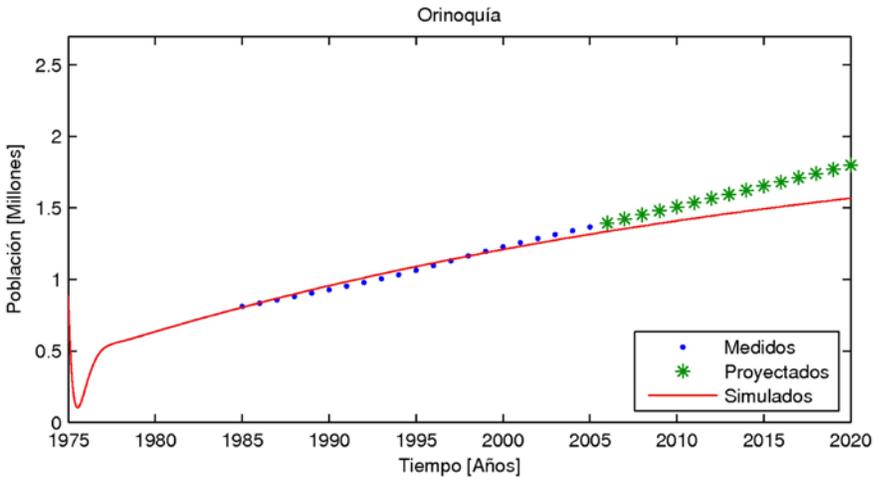


Figura 12. Resultado de simulación para la región Orinoquia con la función J_2 .

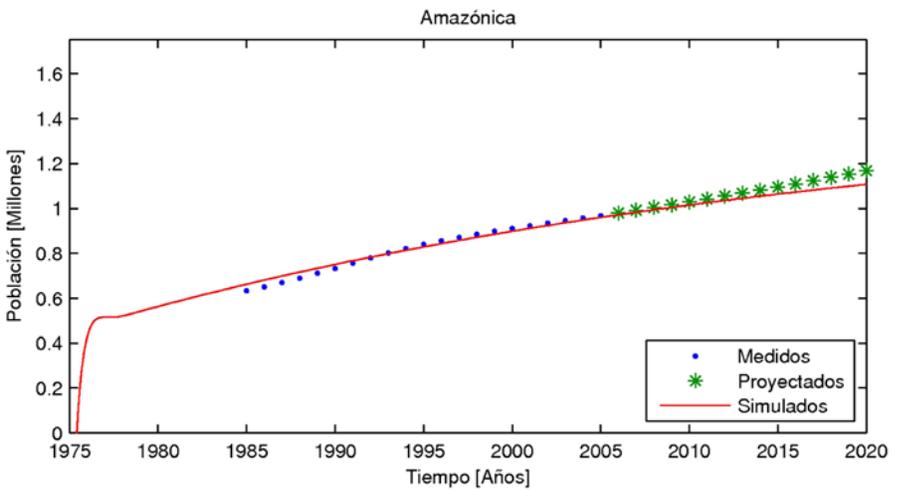


Figura 13. Resultado de simulación para la región Amazonia con la función J_2 .

METODOLOGÍA PARA EL ANÁLISIS ESTADÍSTICO DE RESULTADOS

Cuando presenta variabilidad en los datos de un experimento se suele emplear una prueba estadística de hipótesis con el fin de establecer los re-

sultados del experimento. Con este enfoque se plantea una aseveración y, posteriormente, se realizan pruebas para verificar o rechazar la aseveración

La finalidad del presente análisis consiste en determinar la función objetivo que permite tener un mejor ajuste del modelo propuesto. Para la comparación se emplean los datos obtenidos de las 50 ejecuciones de los algoritmos de optimización para cada función objetivo.

En este caso fueron consideradas las siguientes hipótesis:

- H_0 : Hipótesis nula. Los resultados obtenidos para los grupos experimentales presentaron valores medios iguales.
- H_1 : Hipótesis alternativa. Los resultados obtenidos para los grupos experimentales no presentaron valores medios iguales.

Como es de apreciar, la hipótesis nula en este caso permite establecer de forma estadística si las dos funciones objetivo logran un ajuste similar del modelo. Por su parte, la hipótesis alternativa indica que los modelos obtenidos con las dos funciones objetivos son diferentes desde un punto de vista estadístico.

Adicionalmente, para realizar la prueba de hipótesis se considera un nivel de significancia denominado p -value, es decir que la hipótesis nula se rechaza si el p -value asociado al estadístico de prueba es menor que el nivel de significancia establecido.

Con el fin de aceptar o rechazar las hipótesis existen diferentes pruebas estadísticas, las cuales principalmente se pueden clasificar como paramétricas y no paramétricas.

Las pruebas paramétricas son robustas sin embargo requieren el cumplimiento de consideraciones sobre los datos lo cual no es necesario con las pruebas no paramétricas. En el caso de emplear pruebas paramétricas es necesario comprobar previamente las suposiciones de normalidad y homocedasticidad. La prueba de normalidad consiste en determinar si los datos presentan una distribución normal, mientras que la prueba de homocedasticidad busca establecer si los grupos de datos a comparar presentan la misma varianza.

Cuando se quiere comparar dos grupos independientes, en el caso de cumplirse normalidad y homocedasticidad, se emplea la prueba t ; en caso contrario se utiliza la prueba U de Mann-Whitney.

ANÁLISIS ESTADÍSTICO DE RESULTADOS

Para las pruebas de normalidad, homocedasticidad y comparación entre grupos se tomó un nivel de significancia de 0.05 lo cual corresponde a un error del 5% de aceptación de la hipótesis nula. Para realizar una comparación justa de los resultados obtenidos con las diferentes configuraciones de parámetros establecidos, estos fueron evaluados sobre la misma función objetivo J_1 y J_2 por separado.

Resultados para J_1

Al realizar la prueba de normalidad de Kolmogorov-Smirnov se encontró un p -value de 3.5404×10^{-20} y para la prueba de homocedasticidad de Levene un p -value de 0.0055, de tal forma que se rechazó la hipótesis de normalidad y homocedasticidad, por lo cual fue empleada la prueba U de Mann-Whitney para comparar si los dos grupos de parámetros obtenidos no presentan diferencia significativa. Al realizar esta prueba se encontró un p -value de 0.4087 por lo cual, considerando un nivel de significancia de 0.05 no fue rechazada H_0 .

En la Tabla 3 se pueden apreciar los estadísticos calculados al tomar los resultados obtenidos con J_1 y J_2 y evaluarlos para J_1 .

Tabla 3. Estadísticos al tomar los resultados obtenidos con J_1 y J_2 y evaluarlos para J_1 .

Solución con el método	Mínimo	Media
No ponderado (J_1)	0,0065	8,1756
Ponderado (J_2)	0,0096	25,2750

Resultados para J_2

Al realizar la prueba de normalidad de Kolmogorov-Smirnov se obtuvo un p -value de 1.9838×10^{17} , de tal forma que se rechazó la hipótesis de normalidad, por lo cual se emplea la prueba U de Mann-Whitney con la

cual se encontró un p -value de 0.0421, por lo tanto, considerando un nivel de significancia de 0.05, se rechazó H_0 lo cual indica que los dos grupos de parámetros presentaron diferencia.

En la Tabla 4 se pueden apreciar los estadísticos calculados al tomar los resultados obtenidos con J_1 y J_2 y evaluarlos para J_2 .

Tabla 4. Estadísticos al tomar los resultados obtenidos con J_1 y J_2 y evaluarlos para J_2 .

Solución con el método	Mínimo	Media
No ponderado (J_1)	0,0030	2,1863
Ponderado (J_2)	0,0026	1,4410

CONCLUSIONES

Se apreció que el modelo propuesto logra simular el comportamiento migratorio durante el período de tiempo medido (1985 - 2005). Como es de notar en las simulaciones, el modelo también permite estimar las poblaciones de cada región para años futuros por lo cual puede ser de ayuda para planear la inversión en servicios públicos, escuelas, hospitales y viviendas entre otros.

Considerando que no es posible establecer una población ni un tiempo inicial para los años y datos reportados por el DANE, fue necesario incluir las condiciones iniciales del modelo como variables de optimización y considerar un tiempo de holgura antes de realizar la comparación del resultado del modelo con los datos reales.

La estrategia de optimización empleada, donde se utilizan algoritmos genéticos y minimización basada en gradiente, presenta un buen desempeño para la identificación de parámetros del modelo. Como no se conocían las condiciones iniciales del modelo, los algoritmos genéticos son una buena alternativa para encontrar una solución inicial, la cual, posteriormente, se refina mediante un algoritmo basado en el cálculo de gradientes.

Al emplear una función objetivo sin ponderar se apreció que el modelo se ajusta mejor en las regiones que tienen una mayor población, esto ocurre por ser regiones las más influyentes en la función objetivo. Por el contrario,

se puede apreciar que al emplear la función objetivo ponderada, según el número de individuos en cada región, se tiene un ajuste adecuado de las regiones con un número de individuos bajo. El análisis estadístico mostró que los resultados obtenidos con la función objetivo ponderada son diferentes y presentan un mejor ajuste que los obtenidos con la función no ponderada.

REFERENCIAS

- [1] *Diccionario demográfico multilingüe*, Unión Internacional para el Estudio Científico de la Población, Lieja, Bélgica, 1985.
- [2] G. Díaz, "Aproximaciones metodológicas al estudio de las migraciones internacionales", *UNISCI Discussion Papers*, vol. 15, pp. 157-171, 2007.
- [3] J. Arango, "Las leyes de las migraciones de E.G. Ravenstein, cien años después", *Revista Española de Investigaciones Sociológicas*, vol. 32, pp. 7-26, 1985.
- [4] L. Chávez, V. Kalashnikov, y N. Kalashnycova, "A human migration model with consistent conjectures", *Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*, 2009, pp. 52-55.
- [5] C. Garrocho, "Un modelo de simulación de los flujos de migración interna de México: aplicación empírica de un modelo de interacción espacial", *Estudios Demográficos y Urbanos*, vol. 20, pp. 433-476, 2005.
- [6] M. Guan, "Migration among multicities", *International Journal of Human and Social Sciences*, vol. 5, no. 2, pp. 28-36, 2011.
- [7] P. Wilner, M. Partridge, E. Irwin, "Estimation of a spatial simultaneous equation model of population migration and housing price dynamics", *Regional Science and Urban Economics*, vol. 40, no. 5, pp. 343-352, 2010. DOI:10.1016/j.regsciurbeco.2010.01.002
- [8] A. Martins, "Bilateral mobility in dualistic models", *Economic Modelling*, vol. 25, no. 3, pp. 391-410, 2008.
- [9] H. Booth, R. Hyndman, "Stochastic population forecast using functional data models for mortality, fertility and migration", *International Journal of Forecasting*, vol. 24, no. 3, pp. 323-342, 2008. DOI:10.1016/j.ijforecast.2008.02.009
- [10] C. Christou, "Simulation of Human Migration Based on Swarm Theory", *13th Conference on Information Fusion (FUSION)*, pp. 1-8, 2010.
- [11] L. Marchiori, J. Maystadt, I. Shumacher, "The impact of weather anomalies on migration in Sub-Saharan África", *Journal of Environmental Economics and Management*, vol. 63, no. 3, pp. 355-374, 2012. DOI:10.1016/j.jeem.2012.02.001
- [12] W. Zhang, "Coexistence and separation of two residential groups - An interactional spatial dynamic approach", *Geographical Analysis*, vol. 21, no. 2, pp. 91-102, 2010. DOI: 10.1111/j.1538-4632.1989.tb00881.x

- [13] J. Sagayo, "Desplazamiento forzoso en Colombia: expulsión y movilidad, dos dinámicas que interactúan", Tesis de maestría, Facultad de Ciencias Económicas, Universidad Nacional de Colombia, Bogotá, 2011.
- [14] A. Gemici, "Family migration and labor market outcomes", Ph.D. dissertation, University of Pennsylvania, Pennsylvania, USA, 2006.
- [15] J. Mincer, "Family migration decision", *Journal of Political Economy*, vol. 86, no. 5, pp. 749-773, 1977.
- [16] G. Borjas, S. Bronars, "Immigration and the family", *Journal of Labor Economics*, vol. 9, no. 2, pp. 123-148, 1991.
- [17] H. Foad, "FDI and immigration: a regional analysis", *The Annals of Regional Science*, vol. 49, no. 1, pp. 237-259, 2012. DOI: 10.1007/s00168-011-0438-4
- [18] B. Javorcik, C. Neagu, C. Ozden, M. Spatareanu, "Migrant networks and foreign direct investment", *Journal of Development Economics*, vol. 94, no. 2, pp. 231-241, 2010. DOI:10.1016/j.jdeveco.2010.01.012
- [19] M. Kugler, H. Rapoport, "Skilled emigration, business networks and foreign direct investment", *CESifo Working Paper Series*, no. 1455, 2005.
- [20] T. Weise, "Global optimization algorithms - theory and application", *Self-Published Thomas Weise*, 2009.
- [21] J. Brito, "Aprendizaje Deductivo, Inductivo, y Abductivo con Algoritmos Genéticos", *Revista Ingeniería Informática*, vol. 10, 2004.
- [22] H. Mora, "Optimización no lineal y dinámica", *Universidad Nacional de Colombia*, 2001.
- [23] M. Gestal, D. Rivero, J. R. Rabuñal, J. Dorado, y A. Pazos, "Introducción a los Algoritmos Genéticos y la Programación Genética", *Universidade da Coruña Servizo de Publicacións*, 2010.