

# MODELADO DE FENÓMENOS MIGRATORIOS

TRABAJO FIN DE GRADO  
Curso 2021/2022



**UNIVERSIDAD COMPLUTENSE  
MADRID**

FACULTAD DE CIENCIAS MATEMÁTICAS  
GRADO EN INGENIERÍA MATEMÁTICA

Adriana Barral Díaz  
Tutor: Ana Carpio

Madrid, 29 de junio de 2022

## Resumen

El objetivo de este trabajo es revisar algunos de los modelos matemáticos propuestos para estudiar los fenómenos migratorios. Nos centraremos en aquellos modelos que estudian las migraciones humanas. Para ello, clasificaremos los modelos en cuatro grupos: modelos de Markov, modelos gravitatorios, modelos económicos y modelos basados en la ecuación maestra.

A continuación, trataremos el modelo de radiación para migración humana, un modelo basado en los procesos de emisión y absorción. Se propuso con el fin de evitar las limitaciones de los modelos gravitatorios. Veremos que este modelo permite predecir los flujos migratorios entre dos regiones conociendo la población y el número total de emigrantes de ambas. También veremos el modelo de radiación generalizado, para tener en cuenta otros factores no económicos que influyen en la decisión de migrar.

Estos dos últimos modelos los utilizaremos para realizar una simulación. En ella buscaremos predecir los flujos migratorios de este año entre las provincias españolas y compararemos los resultados obtenidos por ambas.

**Palabras clave:** modelos de Markov, modelos gravitatorios, modelos económicos, modelos basados en la ecuación maestra, modelo de radiación, modelo de radiación generalizado

## Abstract

The aim of this project is to review some of the mathematical models proposed to study migratory phenomena. We will focus on models which study human migrations. To do it, we will classify the models into four groups: Markov models, gravitational models, economic models and models based on the master equation.

Then we will address the radiation model for human migration, a model based on emission and absorption processes. It was proposed in order to avoid the limitations of gravitational models. We will see that this model allows us to predict migratory flows between two regions knowing the population and the total number of emigrants from both. We will also see the generalized radiation model, to take into account other non-economic factors that influence in the decision of migrate.

The last two models will be used to perform a simulation. In it we will seek to predict the migratory flows between the Spanish provinces and we will compare the results obtained by both.

**Keywords:** Markov models, gravitational models, economic models, models based on master equation, radiation model, generalized radiation model

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Revisión de los modelos matemático más utilizados</b>	<b>4</b>
2.1	Modelos de Markov . . . . .	4
2.1.1	Modelo Mover-Stayer . . . . .	5
2.1.2	Modelo de MCFarland . . . . .	6
2.2	Modelos gravitatorios . . . . .	6
2.2.1	Modelo de Zipf . . . . .	7
2.2.2	Modelos de interacción espacial . . . . .	7
2.3	Modelos económicos . . . . .	8
2.3.1	Modelo de Todaro . . . . .	8
2.3.2	Modelo de Cross . . . . .	9
2.4	Modelos basados en la ecuación maestra . . . . .	10
<b>3</b>	<b>Modelo de radiación para migración humana</b>	<b>13</b>
<b>4</b>	<b>Simulación</b>	<b>16</b>
<b>5</b>	<b>Conclusión</b>	<b>32</b>
<b>6</b>	<b>Bibliografía</b>	<b>33</b>

# 1 Introducción

La migración es un fenómeno producido por el desplazamiento de individuos o grupos de individuos de su lugar de origen a otro motivados por diversos factores. Podemos dividirlos en migraciones humanas o migraciones animales. En este trabajo nos centraremos en el primero de los dos.

La migración humana es un fenómeno demográfico que ha estado siempre presente a lo largo de nuestra historia, motivado por causas económicas, políticas o bélicas entre otras. Estos desplazamientos han tenido y tienen consecuencias positivas y negativas tanto en el lugar de origen como en el lugar de destino pero no posee la claridad conceptual de otros fenómenos demográficos como los nacimientos y las muertes, por lo que su estudio es muy complejo. Es por ello por lo que muchos investigadores de diversas disciplinas se han interesado en buscar explicación a este comportamiento obteniéndose como resultado teorías sobre la migración desde diversos puntos de vista.

En primer lugar, realizaremos una revisión de modelos matemáticos que se utilizan para el estudio de los fenómenos migratorios. A estos modelos podemos clasificarlos en cuatro bloques: modelos de Markov, modelos gravitatorios, modelos económicos y modelos basados en la ecuación maestra.

Después nos centraremos en el modelo de radiación, un modelo basado en los procesos de emisión y absorción de radiación y en los mismos principios que los modelos gravitatorios pero que no presenta las limitaciones de estos últimos. Además, trataremos el modelo de radiación generalizado en el que se consideran otros factores, no solo económicos, como la calidad de vida.

Aplicaremos seguidamente el modelo de radiación y el modelo de radiación generalizada para realizar una simulación y ver cómo estiman estos modelos los flujos migratorios en diversos casos y compararemos los resultados obtenidos para ver las similitudes y diferencias observables en ellos.

## 2 Revisión de los modelos matemático más utilizados

En esta sección haremos un repaso de algunos de los modelos utilizados para el estudio de migraciones humanas. Como hemos dicho antes, estos modelos los podemos separar en cuatro bloques: modelos de Markov, modelos gravitatorios, modelos económicos y modelos basados en la ecuación maestra.

### 2.1 Modelos de Markov

Los modelos de Markov reciben su nombre por estar basados en las cadenas de Markov, por lo que antes de hablar de estos modelos repasamos algunos conceptos de las cadenas de Markov que se utilizan en ellos (nos basamos en [9] y [11]).

Una cadena de Markov es un proceso estocástico discreto, es decir, una colección de variables aleatorias  $\{X_n : n \geq 0\}$  que toman valores en un conjunto  $S$  discreto que se llama espacio de estados y que satisfacen la propiedad de Markov

$$P\{X_{n+1} = x_{n+1} | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\},$$

es decir, que la probabilidad de que se produzca un evento depende únicamente del evento inmediatamente anterior.

La función de masa de cada variable aleatoria  $X_n$  se denota por

$$p_i(n) = P\{X_n = i\}, i \geq 0, \forall n \geq 0$$

La probabilidad de transición del estado  $i$  al estado  $j$  se define como la probabilidad de que el proceso se encuentre en el estado  $j$  condicionada a que en el instante anterior estaba en el estado  $i$ , es decir,  $p_{ij}(n) = P\{X_{n+1} = j | X_n = i\}$ . Si estas probabilidades no dependen del instante de tiempo  $n$ , entonces se dice que la cadena de Markov es homogénea. En este caso, la probabilidad de transición se denota  $p_{ij}$ .

A partir de ahora suponemos que trabajamos con cadenas de Markov homogéneas. La matriz de transición de una cadena de Markov es la matriz cuyos elementos son las probabilidades de transición y se denota por  $P = (p_{ij})$ . Esta matriz es estocástica pues verifica que  $p_{ij} \geq 0$  y que cada fila de la matriz es una distribución de probabilidad, es decir,  $\sum_{j \in S} p_{ij} = 1$ .

Las probabilidades de transición de ir del estado  $i$  al estado  $j$  en  $n$  pasos se denota por  $p_{ij}^{(n)} = P\{X_n = j | X_0 = i\}$ . Estas probabilidades son los elementos de la matriz de transición en  $n$  etapas que se denota  $P^{(n)}$ . Esta matriz verifica en general:

$$P^{(n)} = P^n$$

Las Ecuaciones de Chapman-Kolmogorov  $p_{ij}^{(n)} = \sum_{k=1}^{\infty} p_{ik}^{(s)} p_{kj}^{(n-s)}, 0 < s < n$  son la relación entre las probabilidades de transición en  $n$  y  $(n-s)$  pasos.

Por otro lado, dada la matriz de transición en  $n$  pasos  $P^{(n)}$  nos interesa analizar su comportamiento cuando  $n$  tiende a infinito. Si este límite existe, entonces llamamos a la matriz límite

$$H = \lim_{n \rightarrow \infty} P^{(n)} = (h_{ij}),$$

donde  $h_{ij} = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$  verificando que  $0 \leq h_{ij} \leq 1$  y  $\sum_j h_{ij} = 1$ .

Si las filas de esta matriz son idénticas y sabiendo que  $p_{ij}^{(n)} = P\{X_n = j | X_0 = i\}$  entonces la matriz límite es independiente de la distribución inicial y en este caso recibe el nombre de matriz de equilibrio a la que denotamos  $\pi = (\pi_{i,j}) = (h_{ij})$ .

De entre todos los modelos de Markov, nos centraremos en el modelo Mover-Stayer y en el modelo de MCFarland. Para ambos modelos nos basaremos en [4].

### 2.1.1 Modelo Mover-Stayer

El modelo Mover-Stayer fue desarrollado por Blumen, Kogan y MCCarthy en 1955. Parte de dos supuestos:

1. En cada estado  $i$  del espacio de estados  $S$ , la población está dividida en dos grupos. En el primero se encuentran los individuos con probabilidad nula de migrar a los que denominan Stayers, en proporción  $\alpha_i$ , y en el segundo grupo se encuentran los individuos con alguna probabilidad de migrar a los que llaman Movers, en proporción  $1 - \alpha_i$ .
2. La posición del grupo Movers sigue en cada instante de tiempo una cadena de Markov discreta y homogénea en el tiempo.

De acuerdo con estos dos supuestos, el modelo se trata de una mixtura de dos cadenas de Markov independientes, una degenerada con la identidad como matriz de transición y la segunda una cadena de Markov homogénea y discreta en el tiempo con matriz de transición  $M$ .

Teniendo en cuenta todo lo anterior, la matriz de transición en un paso esperada para la población viene dada por:

$$P = \alpha + (I - \alpha)M,$$

donde  $\alpha$  es una matriz diagonal cuyos elementos son la proporción de Stayers en cada estado.

Además, sabiendo que la cadena de Markov que caracteriza al grupo de los Movers es homogénea en el tiempo entonces la matriz de transición en  $n$  pasos esperada para la población se obtiene como:

$$P = \alpha + (I - \alpha)M^n$$

Este modelo ha dado paso a todos los estudios dedicados a la relajación del supuesto markoviano de homogeneidad de la población y es un caso particular del modelo de McFarland que veremos a continuación.

### 2.1.2 Modelo de MCFarland

El modelo de MCFarland (1970) surge como propuesta de relajación a la hipótesis de homogeneidad de la población. En este modelo se define una matriz de transición para cada individuo independiente del tiempo, es decir, cada individuo se mueve independientemente del resto siguiendo una cadena de Markov discreta, finita y homogénea en el tiempo. Suponiendo que tenemos una población compuesta por  $q$  individuos y cada individuo  $c$  tiene matriz de transición  $P_c$ , obtenemos la matriz de transición en un solo paso esperada para la población a partir de la combinación de las matrices de transición de cada individuo:

$$D = \sum_{c=1}^q S_c P_c$$

siendo  $S_c$  la matriz diagonal con  $S_{c,i}$  la proporción de la población que ocupa el estado  $i$  en el instante inicial  $t = 0$  tal que  $\sum_{c=1}^q S_c = I$ .

Considerando todo lo anterior, tenemos entonces  $q$  cadenas de Markov independientes, discretas, finitas y homogéneas en el tiempo. Como vimos antes, la matriz de transición en  $n$  pasos de una cadena de Markov homogénea verifica en general que  $P^{(n)} = P^n$ , por lo que, para cada individuo  $c$ , la matriz de transición en  $n$  pasos es igual a  $P_c^n$ . Así, podemos obtener la matriz de transición en  $n$  pasos esperada para la población de forma similar al de un paso:

$$D(n) = \sum_{c=1}^q S_c P_c^n$$

Teniendo de nuevo en cuenta que tenemos  $q$  cadenas de Markov independientes, discretas, finitas y homogéneas en el tiempo, cada una de ellas da lugar a un matriz de equilibrio. Para el individuo  $c$  denotamos la matriz de equilibrio como  $\pi_c$ . A partir de estas matrices, McFarland define la matriz de equilibrio esperada para la población como:

$$\pi = \sum_{c=1}^q S_c \pi_c$$

A diferencia de la matriz de equilibrio de cada cadena de Markov, las filas de esta matriz no son iguales al depender de  $S_c$ .

El modelo se puede generalizar suponiendo que la población esta agrupada en grupos de acuerdo con alguna característica demográfica donde  $g$  es el número de grupos homogéneos diferenciados con  $1 \leq g \leq q$ . Así, si  $g = q$  tendríamos el modelo de McFarland como caso particular.

## 2.2 Modelos gravitatorios

Los modelos gravitatorios reciben su nombre por estar basados en la ley de gravitación universal de Newton. Estos modelos tienen como objetivo explicar el motivo por el que los movimientos ocurren según lo observado.

### 2.2.1 Modelo de Zipf

Basándonos en [2], este modelo fue propuesto por George K. Zipf (1946). Supone que el número de migrantes que se desplazan entre dos regiones es directamente proporcional al producto de las poblaciones de ambas regiones e inversamente proporcional a la distancia que las separa. De esta forma, sea  $P_i$  la población de la región  $i$ ,  $P_j$  la población de la región  $j$ ,  $D_{ij}$  la distancia entre las dos regiones y  $K$  la fuerza de atracción de la región de destino, entonces el número de migrantes  $M_{ij}$  es:

$$M_{ij} = K \frac{P_i P_j}{D_{ij}}$$

A partir de este modelo se realizan modificaciones en las que consideran otros factores que influyen en las decisiones de migrar, aunque en ellas solo se consideran las relaciones entre las regiones de origen y destino sin tener en cuenta el resto.

### 2.2.2 Modelos de interacción espacial

Estos modelos se utilizan para estudiar las relaciones de los movimientos de personas que tienen lugar en el espacio como consecuencia de un proceso de toma de decisiones. Veamos la Teoría de los movimientos que es un modelo de interacción espacial muy general propuesto por Alonso (1978) y las cuatro familias de modelos de interacción espacial de Wilson (1980) basándonos en [2] y [8].

En la Teoría de los movimientos, los flujos dependen de cinco factores:

1. El primero está relacionado con las características de los lugares de origen y lo denotamos como  $V_i$
2. El segundo está relacionado con las características de los lugares de destino y lo denotamos como  $W_j$
3. El tercero está relacionado con la facilidad de movimiento entre las regiones de origen y de destino. Puede medirse como la inversa del coste de desplazamiento, del tiempo de desplazamiento o de la distancia, siendo este último el más utilizado. Lo denotamos como  $F_{ij}$
4. El cuarto está relacionado con las oportunidades alternativas disponibles en el lugar de origen y lo denotamos como  $A_i$
5. El quinto está relacionado con el grado de competencia existente en el lugar de destino y lo denotamos como  $B_j$

Estos dos últimos factores se definen como:

$$A_i = \sum_k B_k^{\beta_j - 1} W_k F_{ik}$$

$$B_j = \sum_k A_k^{\alpha_i - 1} V_k F_{kj}$$

donde  $\alpha_i$  y  $\beta_j$  son parámetros que indican la respuesta de las regiones a los efectos del sistema y se esperan que estén entre 0 y 1.

Con todo lo anterior, se modelan los flujos entre regiones como

$$T_{ij} = A_i^{\alpha_i-1} V_i B_j^{\beta_j-1} W_j F_{ij}$$

A partir de los flujos podemos definir  $O_i = \sum_j T_{ij} = A_i^{\alpha_i} V_i$  como el flujo total emitido por cada lugar de origen y  $D_j = \sum_i T_{ij} = B_j^{\beta_j} W_j$  como el flujo total recibido en cada lugar de destino.

Posteriormente, Wilson introdujo cuatro familias de modelos de interacción espacial en función de las restricciones:

1. Modelos no restrictivos o modelo de interacción tradicional (no se conocen ni los  $O_i$  ni los  $D_j$ ):  $T_{ij} = V_i W_j F_{ij}$
2. Modelos con restricción en el origen (conocemos  $O_i$ ):  $T_{ij} = A_i^{-1} O_i W_j F_{ij}$
3. Modelos con restricción en el destino (conocemos  $D_j$ ):  $T_{ij} = B_j^{-1} V_i D_j F_{ij}$
4. Modelos con doble restricción (conocemos  $O_i$  y  $D_j$ ):  $T_{ij} = A_i B_j O_i D_j F_{ij}$

Estos modelos se pueden obtener como caso particular del modelo de la Teoría de los movimientos tomando como valores de  $\alpha$  y  $\beta$  el 0 y 1:

1. Si  $\alpha = \beta = 1$  tenemos el modelo de interacción tradicional.
2. Si  $\alpha = 0$  y  $\beta = 1$  tenemos el modelo con restricción en el origen.
3. Si  $\alpha = 1$  y  $\beta = 0$  tenemos el modelo con restricción en el destino.
4. Si  $\alpha = \beta = 0$  tenemos el modelo con doble restricción.

## 2.3 Modelos económicos

Los modelos económicos son aquellos que intentan explicar las migraciones desde un punto de vista fundamentalmente económico. Nos centraremos en el modelo de Todaro y en el modelo de Cross basándonos en [6] y [1] respectivamente.

### 2.3.1 Modelo de Todaro

Según este modelo, la decisión de migrar de áreas rurales a urbanas está relacionada con dos variables: la diferencia de ingreso real entre la zona rural y la urbana y la probabilidad de obtener un trabajo en la zona urbana. Además, el modelo parte de cuatro supuestos:

1. El cambio porcentual en la fuerza laboral como resultado de la migración durante cualquier periodo se rige por  $\frac{\dot{S}}{S}(t) = F \left[ \frac{V_u(t) - V_r(t)}{V_r(t)} \right]$ ,  $F' > 0$ , donde  $\dot{S}$  es el flujo de mano de obra procedente de la zona rural hacia la zona urbana,  $S$  es la cantidad de mano de obra existente en la zona urbana,  $V_u(t)$  y  $V_r(t)$  es el valor presente de renta real esperada a lo largo del horizonte temporal en la zona urbana y rural en el período  $t$  respectivamente.
2. El horizonte de planificación para cada trabajador es idéntico.

3. Los costes fijos de la migración son idénticos para todos los trabajadores.
4. El factor de descuento es constante durante el horizonte de planificación e idéntico para todos los migrantes.

Dados estos supuestos el modelo se formula de la siguiente manera:

$$V_r(0) = \int_{t=0}^n Y_r(t)e^{-rt} dt$$

y

$$V_u(0) = \int_{t=0}^n p(t)Y_u(t)e^{-rt} dt - C(0),$$

donde  $Y_r(t)$  es el ingreso rural real neto esperado en el periodo  $t$ ,  $Y_u(t)$  es el ingreso urbano real neto esperado en el periodo  $t$ ,  $r$  es el factor de descuento que refleja el grado de preferencia temporal de consumo del trabajador rural,  $C(0)$  es el coste fijo inicial de la migración y reubicación en la zona urbana y  $p(t)$  es la probabilidad de tener trabajo en la zona urbana en el período  $t$ . Esta probabilidad está relacionada con la probabilidad de ser seleccionado entre los trabajadores del grupo en el periodo  $t$  ( $\pi(t)$ ) de la siguiente forma:

$$p(0) = \pi(0),$$

$$p(t) = p(t-1) + [1 - p(t-1)]\pi(t) = \pi(0) + \sum_{i=1}^t \pi(i) \prod_{j=0}^{i-1} (1 - \pi(j))$$

Se define  $\pi(t) = \frac{\gamma N(t)}{S(t) - N(t)}$ , donde  $\gamma = \lambda - p$ , siendo  $\lambda$  y  $p$  la tasa de crecimiento de producción industrial y de productividad laboral respectivamente, y  $N(t) = N_0 e^{\gamma t}$  es el empleo total en la zona urbana en el periodo  $t$ .

### 2.3.2 Modelo de Cross

En este modelo solo se consideran dos zonas, la zona rural donde el salario es  $W_r$  y la zona urbana donde habrá dos trabajos, uno en el sector moderno donde el salario es  $W_u$  y otro trabajo en el que el salario es  $W_0$ . Los salarios verifican que  $W_u > W_r > W_0$ . Además, la probabilidad de encontrar un trabajo en el sector moderno es  $q_i$  y depende de factores como la edad del trabajador o el tiempo de permanencia en la zona urbana. El trabajador individual  $i$  se encontrará en el sector urbano durante el periodo  $t$  con una probabilidad  $P_t^i$  y en el sector rural con una probabilidad  $1 - P_t^i$ . Esta probabilidad se obtendrá de la experiencia.

Sea  $\alpha_i(W)$  la tasa de aprendizaje en función de la magnitud del salario. Esta función verifica que:

1.  $0 < \alpha_i(W) < 1$
2.  $\alpha_i'(W) > 0$
3.  $\alpha_i''(W) < 0$

Conociendo la tasa de aprendizaje obtenemos la probabilidad de estar en la zona urbana en el periodo  $t+1$  considerando las tres posibles situaciones en la que se puede encontrar el trabajador:

1. Si el trabajador  $i$  está en la zona urbana y encuentra trabajo en el sector moderno, entonces  $P_{t+1}^i = P_t^i + \alpha_i(W_u)(1 - P_t^i)$
2. Si el trabajador  $i$  está en la zona urbana y no encuentra trabajo en el sector moderno, entonces  $P_{t+1}^i = P_t^i + \alpha_i(W_0)(1 - P_t^i)$
3. Si el trabajador  $i$  está en la zona rural entonces  $P_{t+1}^i = P_t^i(1 - \alpha_i(W_r))$

Cada una de las situaciones se da con probabilidad  $P_t^i q_i$ ,  $P_t^i(1 - q_i)$  y  $(1 - P_t^i)$  respectivamente.

Como consecuencia de estas tres relaciones tenemos que el valor esperado de  $P_{t+1}^i$  viene dado por:

$$E[P_{t+1}^i] = P_t^i + P_t^i(1 - P_t^i)(q_i \alpha_i(W_u) + (1 - q_i) \alpha_i(W_0) - \alpha_i(W_r))$$

Suponiendo que  $N$  es el total de la población, tenemos que la población esperada para el periodo  $t$  en la zona urbana será  $S(t) = \sum_{i=1}^N P_t^i$  y para el periodo  $t+1$  será  $S(t+1) = \sum_{i=1}^N E[P_{t+1}^i]$

La migración esperada es la diferencia entre la población esperada en el periodo  $t+1$  y la población esperada en el periodo  $t$ , es decir,

$$M(t) = S(t+1) - S(t) = \sum_{i=1}^N P_t^i + P_t^i(1 - P_t^i)(q_i \alpha_i(W_u) + (1 - q_i) \alpha_i(W_0) - \alpha_i(W_r)) - \sum_{i=1}^N P_t^i = \sum_{i=1}^N P_t^i(1 - P_t^i)(q_i \alpha_i(W_u) + (1 - q_i) \alpha_i(W_0) - \alpha_i(W_r))$$

Siempre que el salario urbano esperado sea mayor que el salario rural entonces  $M(t) \geq 0$ .

## 2.4 Modelos basados en la ecuación maestra

Estos modelos se emplean para describir fenómenos, aunque no estén en equilibrio. Para ello se utiliza la ecuación maestra. La ecuación maestra migratoria es una ecuación diferencial estocástica que nos permite estudiar cómo evoluciona la distribución de probabilidad de la población del sistema. Estudiaremos estos modelos basándonos en [2] y [3].

Sean  $n = (n_1, n_2, \dots, n_L)^t$  la distribución poblacional en un tiempo dado, donde  $n_i$  es la población en la región  $i$ ,  $P(n; t)$  la probabilidad de que se dé el estado  $n$  en el período  $t$  y  $W(n/m; t)$  la tasa de transición desde el estado  $m$  al estado  $n$  por unidad de tiempo definida como:

$$W(n/m; t) = \frac{\partial}{\partial \Delta t} P(n; t + \Delta t/m; t)|_{\Delta t=0}$$

Entonces, se define la ecuación maestra como:

$$\frac{d}{dt} P(n; t) = \sum_{m \neq n} [W(n/m; t) P(m; t) - W(m/n; t) P(n; t)]$$

Los cambios que se producen en el sistema se deben a las migraciones, a las que dividimos en migración interna (Mint) y migración externa (Mext), y al crecimiento natural (Cn) de la población. Teniendo esto en cuenta, podemos descomponer la ecuación maestra migratoria como sigue:

$$\frac{d}{dt}P(n; t) = \left[ \frac{d}{dt}P(n; t) \right]_{Mint} + \left[ \frac{d}{dt}P(n; t) \right]_{Cn} + \left[ \frac{d}{dt}P(n; t) \right]_{Mext}$$

Denotamos la tasa de transición de la región  $i$  a la región  $j$  por  $W_{ij}$ . En migraciones interiores, migraciones exteriores y crecimiento natural, estas tasas se obtienen de distinta manera. Antes de ver cada una por separado definimos la tasa de transición individual y las funciones de utilidad.

La tasa de transición individual por unidad de tiempo desde el estado  $i$  al estado  $j$  se define como  $q_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{ij}(t, \Delta t)}{\Delta t}$ , donde  $P_{ij}(t, \Delta t)$  es la probabilidad de transición de la región  $i$  a la región  $j$  durante el intervalo de tiempo  $[t, t + \Delta t]$ . Para migración exterior, se denota como  $q_{Ej}(t)$  cuando la transición se produce desde la región exterior a la región  $j$  ó  $q_{iE}(t)$  cuando vamos de la región  $i$  a la región exterior.

Estas tasas se pueden expresar en función de las llamadas funciones de utilidad  $u_i$ , las cuales caracterizan la atracción de cada región, cumpliendo tres condiciones:

1. Ser positivas
2. Ser  $q_{ij} > q_{ji}$  si  $u_j > u_i$
3. Ser una función monótona no decreciente de  $(u_j - u_i)$

Hay tres expresiones que se utilizan:

1.  $q_{ij} = \exp(u_j - u_i)$
2.  $q_{ij} = \exp\left(\frac{u_j}{u_i}\right)$
3.  $q_{ij} = \frac{u_j}{u_i}$

Las tasas de transición individual pueden incluir factores que reflejan efectos económicos, sociales o de distancia geográfica. En ese caso, sus expresiones se multiplican por un factor de movilidad  $\nu_{ij}(t)$  que normalmente es simétrico. El factor de movilidad se puede descomponer como  $\nu_{ij}(t) = \nu_0(t)f_{ij}$ , donde  $\nu_0(t)$  es el factor de movilidad global y  $f_{ij}$  es un término que depende de los estados origen y destino y es posible que también del tiempo. Este último término se puede relacionar con variables explicativas como la distancia económica o geográfica.

Por otro lado, a las funciones de utilidad las podemos descomponer en la suma de dos términos, uno dependiente del tamaño ( $s_i$ ) y otro que indica las preferencias independientes del tamaño ( $\delta_i$ ). Entonces:

$$u_i(n_i; t) = s_i(n_i; t) + \delta_i(t)$$

Existen varias posibilidades sobre la forma de  $s_i(n_i; t)$  quedando la función de utilidad como:

1.  $u_i(n_i; t) = kn_i(t) + \sigma n_i^2(t) + \delta_i(t)$
2.  $u_i(n_i; t) = k \ln(n_i(t)) + \delta_i(t)$
3.  $u_i(n_i; t) = \delta_i(t)$

A los parámetros  $\{\delta, k, \sigma\}$  se les conoce como parámetros de tendencia del sistema, siendo  $k > 0$  el parámetro que representa la tendencia de la población a concentrarse y  $\sigma$  el parámetro que recoge todos los factores que dependen de las interacciones entre individuos de la región conocido como parámetro de saturación.

Vemos ahora las expresiones de la tasa de transición de la región i a la región j en los distintos casos:

- En migraciones interiores podemos expresar la tasa de transición como  $W_{ij}(n) \equiv W_{ij}(n_i, n_j) = q_{ij}(t)n_i$ . Por lo visto antes, para estas tasas de transición se han propuesto tres modelos:

1.  $W_{ij}(n_i, n_j) = n_i \nu_{ij}(t) \exp(u_j(n_{j+1}) - u_i(n_i))$
2.  $W_{ij}(n_i, n_j) = n_i \nu_{ij}(t) \exp\left(\frac{u_j(n_{j+1})}{u_i(n_i)}\right)$
3.  $W_{ij}(n_i, n_j) = n_i \nu_{ij}(t) \frac{u_j(n_{j+1})}{u_i(n_i)}$

Observamos que la función de utilidad de la región i dependen de la población que hay en esa región antes de la migración y la función de utilidad de la región j depende de la población que hay en esa región después de esa migración.

- En migraciones exteriores se puede expresar como:
  - $W_{iE} = q_{iE}(t)n_i^*$  donde  $n_i^*$  es el número de personas de la región i que hay en la región exterior.
  - $W_{Ei} = q_{Ei}(t)n_i^{**}$  donde  $n_i^{**}$  es el número de personas de la región exterior que hay en la región i.
- En el crecimiento natural distinguimos la tasa de transición de nacimientos  $W_i^+(n) = \beta_i(t)n_i$  y la tasa de transición de muertes  $W_i^-(n) = \mu_i(t)n_i$ , siendo  $\beta_i$  y  $\mu_i$  las tasas de nacimiento y muerte por unidad de tiempo respectivamente.

### 3 Modelo de radiación para migración humana

Los modelos gravitatorios vistos anteriormente presentan una serie de limitaciones a pesar de su uso generalizado. Algunas de estas limitaciones son:

1. No puede predecir la movilidad en regiones donde no tenemos datos.
2. Carece de orientación teórica por lo que se usan hasta nueve parámetros para ajustar los datos empíricos.
3. Predice que el número de desplazamientos aumenta sin límite a medida que aumenta la población de destino, sin tener en cuenta que el número de migrantes no puede superar al de la población de su lugar de origen.

Motivados por estas limitaciones proponen (en [7]) un modelo denominado modelo de radiación basado en los mismos principios que los modelos gravitatorios, pero sin estas limitaciones. Este modelo se basa en los procesos de emisión y absorción de radiación, de ahí su nombre.

En los procesos de emisión y absorción, las partículas se emiten desde una ubicación dada y tienen una probabilidad de ser absorbidas por las ubicaciones que las rodean. Podemos definir estos procesos siguiendo dos pasos:

1. Suponemos que en la ubicación  $i$  hay un total de  $m_i$  partículas. De esta ubicación se emiten algunas de estas partículas ( $X$ ), a las que asociamos un valor que representa el umbral de absorción. A este valor lo denotamos por  $z_X^i$  y se define como el máximo número obtenido después de  $m_i$  extracciones aleatorias de una distribución  $p(z)$ . Así, las partículas emitidas por las localizaciones más pobladas tendrán un umbral de absorción mayor en promedio que las partículas emitidas por localizaciones menos pobladas.
2. Tomamos ahora una ubicación vecina  $j$ , donde hay un total de  $m_j$  partículas y asociamos a las mismas partículas emitidas por la ubicación  $i$  otro valor que representa la absorbencia que tiene  $j$ . A este valor lo denotamos  $z_X^j$  y lo definimos como el máximo número obtenido después de  $m_j$  extracciones aleatorias de la misma distribución que antes,  $p(z)$ . De igual forma que el umbral de absorción, la absorbencia será mayor en promedio en aquellas zonas donde haya más población. Además, las partículas serán absorbidas por las localizaciones más próximas.

Siguiendo estos mismos pasos y suponiendo que las ubicaciones son las zonas entre las que se producen las migraciones, que el número total de partículas es la población de estas y que las partículas emitidas son las personas que piensan desplazarse en busca de mejores oportunidades, obtenemos los flujos de migración entre las distintas zonas. Además, teniendo en cuenta todo lo anterior, las personas migraran a zonas próximas a su lugar de origen cuya absorbencia sea mayor que el umbral de absorción.

Podemos predecir los flujos de migración entre dos zonas calculando la probabilidad de que una persona de la zona  $i$  con población  $m_i$ , se desplace a la zona

j con población  $m_j$ , conociendo la población  $s_{ij}$ , situada en el círculo de centro la zona i y radio la distancia entre i y j sin contar con la población de ambas:

$$p_{ij} = P(1|m_i, m_j, s_{ij}) = \int_0^\infty dz P_{m_i}(z) P_{s_{ij}}(< z) P_{m_j}(> z)$$

donde:

1.  $P_{m_i}(z) = \frac{dP_{m_i}(< z)}{dz} = m_i p(< z)^{m_i-1} \frac{dp(< z)}{dz}$  es la probabilidad de que  $z_X^i$  sea igual a z.
2.  $P_{s_{ij}}(< z) = p(< z)^{s_{ij}}$  es la probabilidad de que el máximo valor después de  $s_{ij}$  extracciones aleatorias de  $p(z)$  sea menor que z.
3.  $P_{m_j}(> z) = 1 - p(< z)^{m_j}$  es la probabilidad de que  $z_X^j$  sea mayor o igual que z.

Entonces

$$\begin{aligned} p_{ij} &= P(1|m_i, m_j, s_{ij}) = m_i \int_0^\infty dz p(< z)^{m_i-1} \frac{dp(< z)}{dz} p(< z)^{s_{ij}} (1 - p(< z)^{m_j}) = \\ &= m_i \int_0^\infty dz \frac{dp(< z)}{dz} (p(< z)^{m_i-1+s_{ij}} - p(< z)^{m_i+s_{ij}+m_j-1}) = \\ &= m_i \left( \frac{1}{m_i + s_{ij}} - \frac{1}{m_i + s_{ij} + m_j} \right) = \frac{m_i m_j}{(m_i + s_{ij})(m_i + s_{ij} + m_j)} \end{aligned}$$

Vemos que la probabilidad es independiente de la distribución  $p(z)$  y es invariante al cambio de escala de la población con el mismo factor. Además,

$$\sum_{j \neq i} p_{ij} = \sum_{j \neq i} \frac{m_i m_j}{(m_i + s_{ij})(m_i + s_{ij} + m_j)} = 1$$

Sea  $T_i$  el número total de personas que se marchan de la región i, la probabilidad de que exactamente  $T_{ij}$  personas se desplacen a la región j es:

$$P(T_{ij}|m_i, m_j, s_{ij}) = \frac{T_i!}{T_{ij}!(T_i - T_{ij})!} p_{ij}^{T_{ij}} (1 - p_{ij})^{T_i - T_{ij}}$$

que es una distribución binomial con media

$$\langle T_{ij} \rangle = T_i p_{ij} = T_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + s_{ij} + m_j)}$$

Esta es la ecuación fundamental del modelo de radiación. Observamos que esta ecuación es independiente de la distancia y además es libre de parámetros.

Este modelo no es válido para países en desarrollo y tampoco considera a las personas que migran por factores no económicos. Por esta razón, en [15] proponen un modelo generalizado del modelo de radiación para migración humana considerando un índice de urbanización U en lugar de usar la población. En este modelo se busca estudiar las migraciones motivadas por factores que no

sean solo económicos, por lo que, considerando k factores se define el índice de urbanización como:

$$U = \sum_k w_k f_k$$

siendo  $w_k$  el peso del factor k ( $f_k$ ).

El índice de urbanización en el modelo de radiación generalizada se utiliza de forma análoga a la población en el modelo de radiación, quedando la ecuación fundamental del modelo generalizado de la siguiente forma:

$$\langle T_{ij} \rangle = T_i \frac{U_i U_j}{(U_i + \nu_{ij})(U_i + \nu_{ij} + U_j)}$$

donde  $U_i$  y  $U_j$  son los índices de urbanización de las zonas i y j, y  $\nu_{ij}$  es el índice de urbanización total en el círculo de centro la región i y radio la distancia entre la región i y j sin contar con los índices de urbanización de ambas.

Debemos tener en cuenta que los factores que consideremos para calcular el índice de urbanización no estarán en las mismas unidades, por lo que para poder compararlas es necesario normalizarlas. Vemos a continuación tres de los métodos de normalización que podemos utilizar (de [15] y [16]):

1. El método Min-max toma el valor mínimo y máximo y los desplaza a los valores 0 y 1 respectivamente utilizando la ecuación  $x' = \frac{x - \min}{\max - \min}$
2. El método puntuación z transforma los datos a una distribución de media 0 y desviación típica 1 a través de la ecuación  $x' = \frac{x - \bar{x}}{\sigma}$ , donde  $\bar{x}$  y  $\sigma$  son la media y la desviación estándar de los datos respectivamente.
3. Un tercer método posible es utilizar una función logística o sigmoidea que transforma los datos en valores entre 0 y 1. Esta función es  $x' = \frac{1}{1 + e^{-t}}$ , donde  $t = \frac{x - \bar{x}}{\sigma}$ , es decir, primero normalizamos los datos con el método de puntuación z y después utilizamos la función logística.

Con respecto a los pesos asociados a los factores, estos pueden ser cualquier número real. Dependiendo del valor que tome, podemos interpretar los pesos de tres formas:

1. Si  $w_k > 0$  implica que el factor  $f_k$  impulsa a las personas a migrar.
2. Si  $w_k = 0$  entonces el factor  $f_k$  no influye en la decisión de migrar.
3. Si  $w_k < 0$  implica que el factor  $f_k$  no impulsa a las personas a migrar.

## 4 Simulación

En esta sección realizaremos una predicción de las migraciones entre las distintas provincias españolas de este año utilizando el modelo de radiación visto en la sección anterior y después realizaremos lo mismo con el modelo de radiación generalizado también visto en la sección anterior.

Para predecir los flujos migratorios con el modelo de radiación, necesitamos la población total de cada provincia, la población situada en los círculos de centro la región de origen y radio la distancia entre la región de origen y destino, sin contar con la población de ambas regiones y el número total de personas que migran de cada provincia.

La población total de cada provincia en el año 2021 la obtenemos del Instituto Nacional de Estadística (INE [10]).

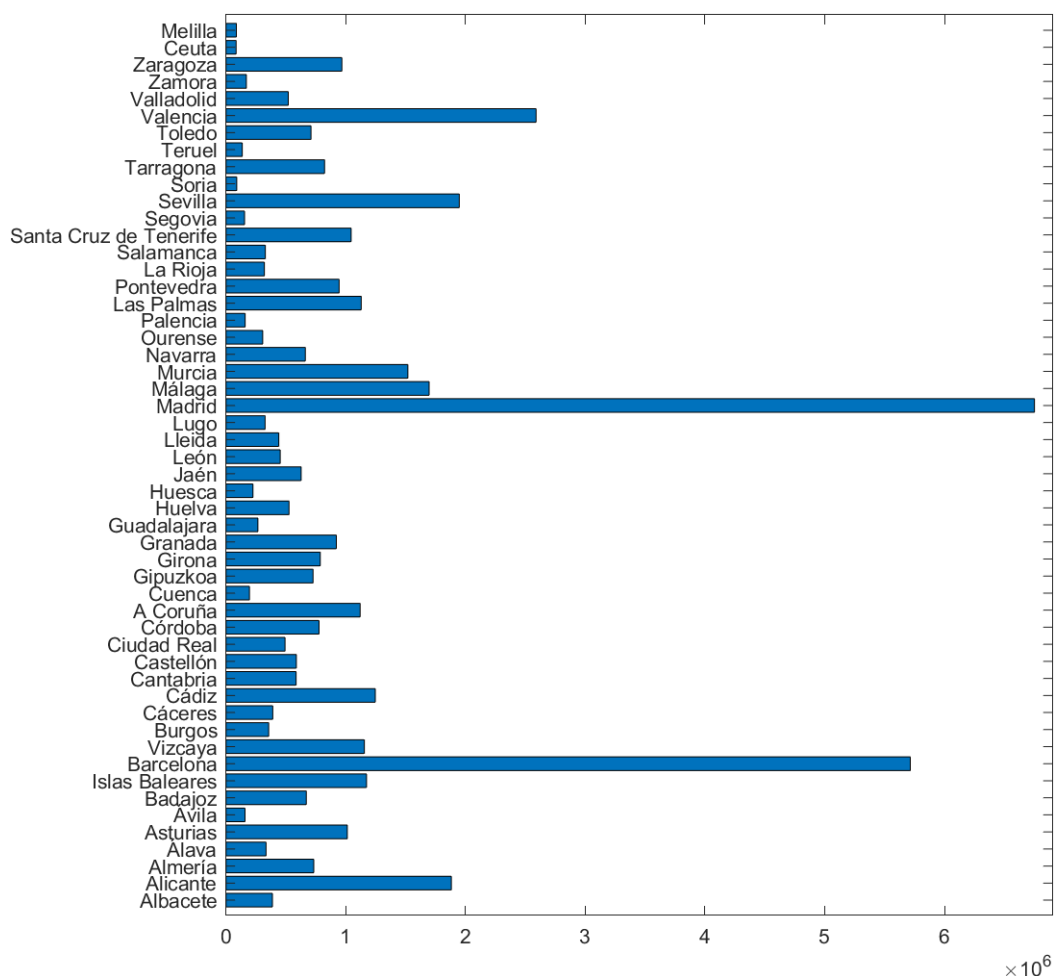


Figura 1: Histograma con la población total en cada provincia española en el año 2021.

Estas poblaciones las usaremos para calcular la población situada en cada círculo, pero antes necesitamos la distancia entre provincias. Estas distancias no son fáciles de conseguir directamente, por lo que para calcularlas tomamos las coordenadas de latitud y longitud de cada provincia de Google Maps y utilizamos la fórmula de Haversine.

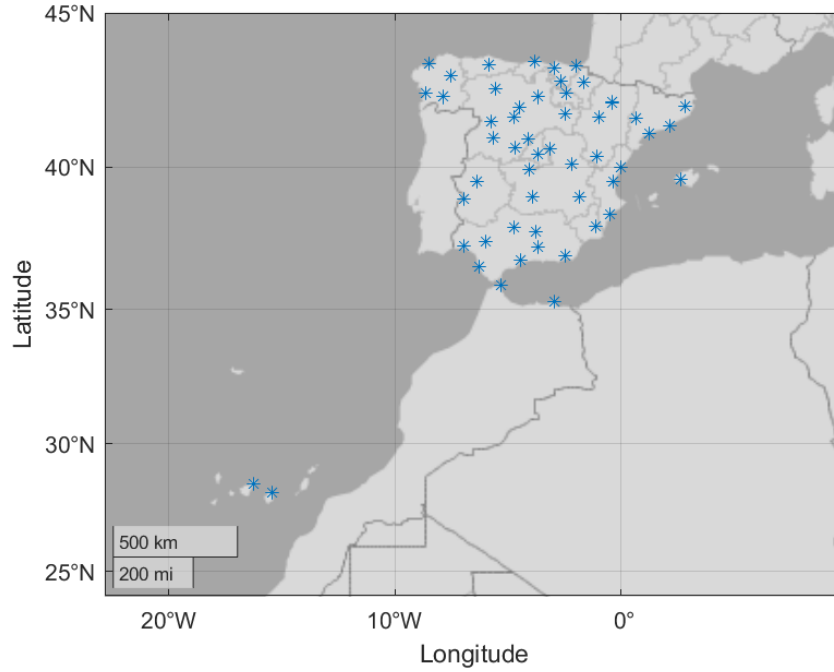


Figura 2: Puntos que representan las provincias en el mapa según las coordenadas tomadas.

La fórmula de Haversine ([13]) es una ecuación importante en la navegación que proporciona las distancias entre dos puntos en una esfera a partir de las longitudes y latitudes:

$$d = 2r \arcsin \left( \sqrt{\left( \sin \frac{\phi_2 - \phi_1}{2} \right)^2 + \cos(\phi_1) \cos(\phi_2) \left( \sin \frac{\psi_2 - \psi_1}{2} \right)^2} \right),$$

donde  $(\phi_1, \psi_1)$  y  $(\phi_2, \psi_2)$  son la latitud y longitud de los puntos respectivamente, y  $r$  es el radio de la Tierra (6371 km).

Una vez calculadas todas las distancias entre provincias vemos la población que hay en cada círculo. Para ello, tomamos la distancia entre la provincia de origen y de destino y vemos cuales de las provincias (sin tener en cuenta las de origen y destino) se sitúan a igual o menor distancia respecto a la provincia de origen. De esta forma, la población que hay dentro del círculo será igual a la suma de la población total de todas las provincias que verifiquen lo anterior.

Conociendo la población total de las 52 provincias y la población de los círculos, calculamos la probabilidad de que una persona se desplace de una provincia a otra. A partir de estas probabilidades podemos comprobar que, como vimos en teoría, las localizaciones más pobladas tendrán un umbral de absorción y una absorbencia mayor en promedio que las localizaciones menos pobladas y las personas migrarán a regiones más próximas. Para ello, decidimos tomar las cuatro provincias más pobladas que son Madrid, Barcelona, Valencia y Alicante, junto con tres de las provincias colindantes con Madrid (Toledo, Ávila y Guadalajara) cuyos datos son:

Provincia	Población total	Latitud	Longitud
Madrid	6.751.251	40,43563	-3,68118
Barcelona	5.714.730	41,37695	2,17668
Valencia	2.589.312	39,48613	-0,36867
Alicante	1.881.762	3,837615	-0,52
Toledo	709.403	39,92455	-4,05602
Ávila	158.421	40,65695	-4,68185
Guadalajara	265.588	40,62447	-3,17138

Tabla 1: Datos de las provincias obtenidos del INE (Población total) y de Google Maps (Latitud y longitud).

Para estas siete provincias tenemos las siguientes matrices con la distancia entre ellas (en km), la población total en los círculos (teniendo en cuenta el resto de las provincias) y las probabilidades de que una persona se traslade de una provincia a otra respectivamente:

destino origen	Madrid	Barcelona	Valencia	Alicante	Toledo	Ávila	Guadalajara
Madrid	0	503,186	301,393	355,214	65,143	88,059	47,9302
Barcelona	503,186	0	301,001	405,279	549,904	580,803	456,457
Valencia	301,393	301,001	0	124,116	319,161	389,367	270,025
Alicante	355,214	405,279	124,116	0	350,125	437,834	337,977
Toledo	65,143	549,904	319,161	350,125	0	97,209	108,117
Ávila	88,059	580,803	389,367	437,834	97,209	0	127,497
Guadalajara	47,9302	456,457	270,025	337,977	108,117	127,497	0

Tabla 2: Distancias entre provincias calculadas con la fórmula de Haversine usando Matlab.

origen \ destino	Madrid	Barcelona	Valencia	Alicante	Toledo	Ávila	Guadalajara
Madrid	0	29496666	7981657	16697137	265588	1128654	0
Barcelona	15612221	0	5134965	9715906	23101642	24489529	12317758
Valencia	14310915	8596185	0	721609	21643504	26554819	8106333
Alicante	11154606	21871737	1518486	0	10445203	28053141	9357306
Toledo	0	36828426	13862002	22257500	0	6751251	7402263
Ávila	153663	37293147	22905996	32088080	7232252	0	8461016
Guadalajara	0	33266259	12222077	20994371	7100430	7809833	0

Tabla 3: Población en los círculos centrados en la provincia de origen y de radio la distancia entre la provincia de origen y de destino de la tabla anterior.

origen \ destino	Madrid	Barcelona	Valencia	Alicante	Toledo	Ávila	Guadalajara
Madrid	0	0.025365	0.068498	0.021389	0.088342	0.016885	0.03785
Barcelona	0.064429	0	0.101483	0.040255	0.004765	0.000987	0.004599
Valencia	0.043734	0.078277	0	0.283406	0.003039	0.00048	0.005866
Alicante	0.049249	0.015363	0.239245	0	0.008307	0.000331	0.003865
Toledo	0.904914	0.002497	0.007346	0.002339	0	0.001977	0.002773
Ávila	0.485194	0.00056	0.000693	0.000271	0.001877	0	0.000549
Guadalajara	0.96215	0.001153	0.003653	0.001016	0.003167	0.000633	0

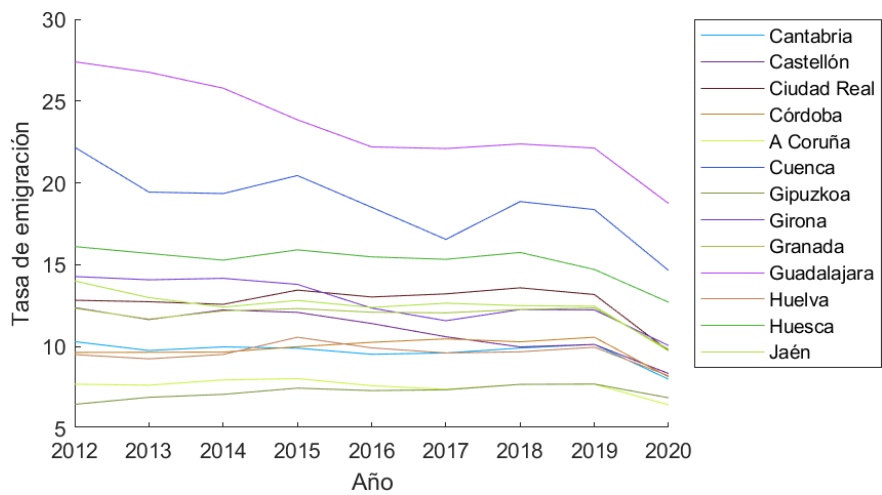
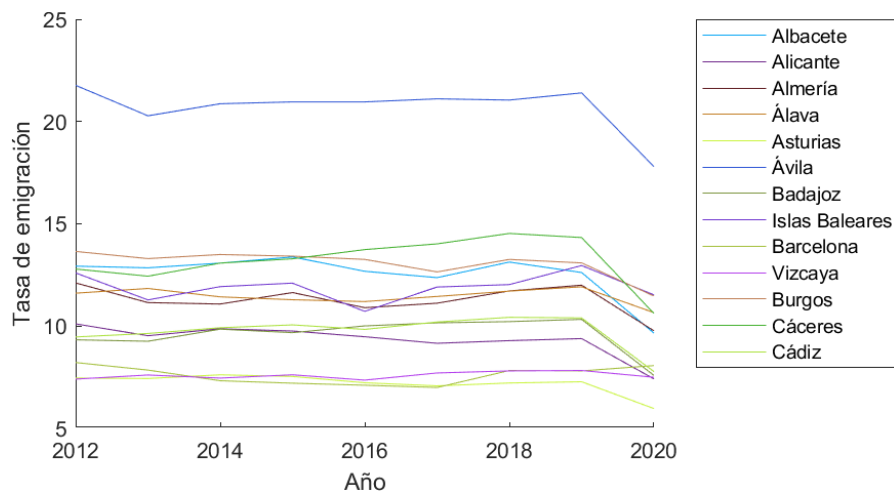
Tabla 4: Probabilidades de que una persona de la provincia de origen migre a la provincia de destino.

En primer lugar, nos fijamos en la columna correspondiente a Madrid en la matriz de probabilidades. Podemos observar en ella que la probabilidad de que una persona de Toledo, Ávila o Guadalajara migre a Madrid es mayor que la probabilidad de que una persona de Barcelona, Valencia o Alicante se traslade a Madrid. Como hemos visto en teoría, las personas migrarán a zonas próximas a su lugar de origen cuya absorbencia sea mayor que el umbral de absorción, por lo tanto, esta diferencia en las probabilidades se debe a la distancia entre provincias y a que la absorbencia de Madrid es mayor que el umbral de absorción de Toledo, Ávila y Guadalajara.

Ahora nos fijamos en la fila correspondiente a Madrid de la misma matriz, en ella están las probabilidades de que una persona de Madrid se traslade al resto de provincias. Como podemos ver, estas probabilidades son muy pequeñas, incluso con las provincias colindantes, lo que se debe a que el umbral de absorción de Madrid es mayor que la absorbencia del resto de provincias.

El resto de las probabilidades son pequeñas debido a que la distancia entre provincias es muy grande, a excepción de las probabilidades de que una persona migre desde Alicante a Valencia y viceversa, pues ambas son provincias colindantes.

Por otro lado, volviendo de nuevo a las 52 provincias, nos faltaría el número total de personas que se van de cada una. Este valor es difícil de estimar pues hay muchos factores que pueden influir en la toma de esta decisión. Debido a eso hemos decidido tomar las tasas de emigración interprovincial por cada mil habitantes del 2012 al 2020 del INE ([14]).



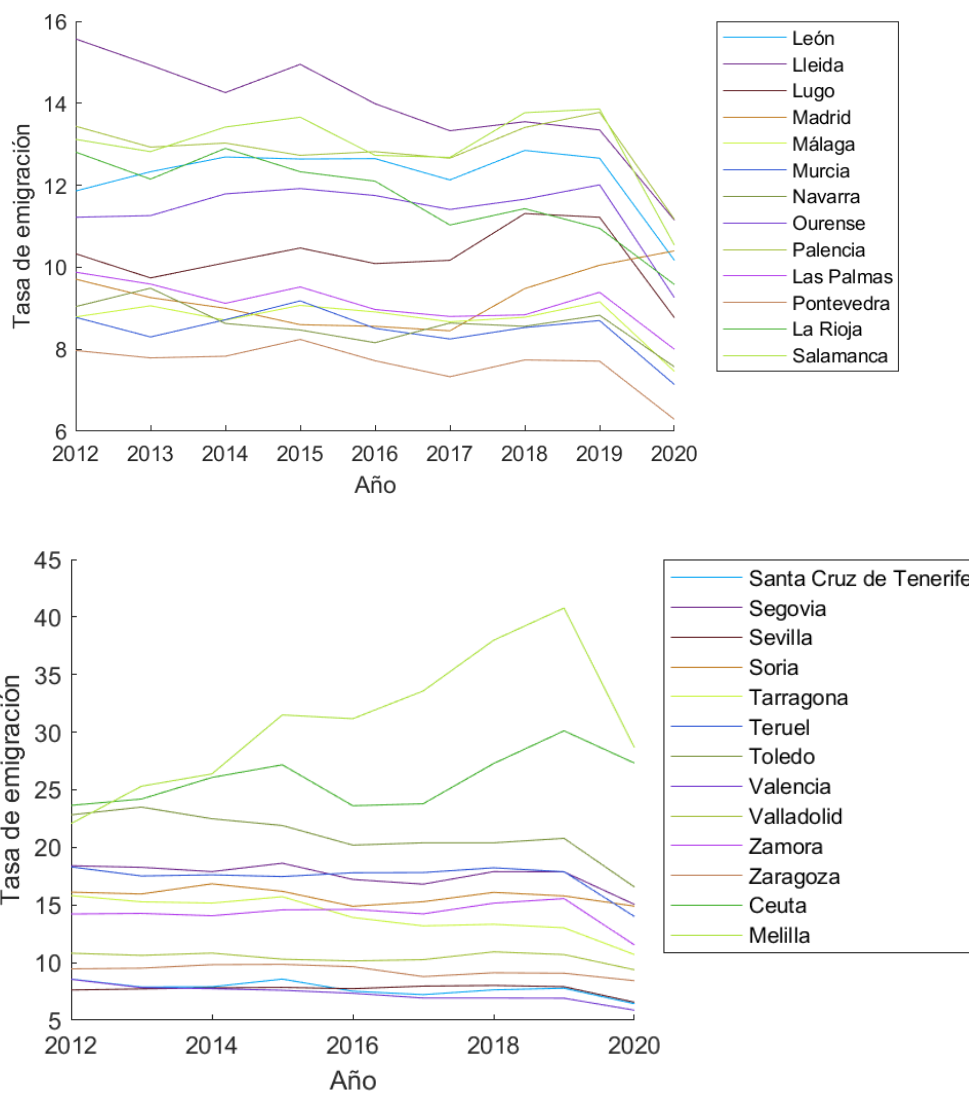


Figura 3: Representación de las tasas de emigración de cada provincia por año.

Observamos que las tasas de emigración interprovincial del 2012 al 2019 son más o menos próximos entre sí. Sin embargo, en la mayoría de los casos las tasas de emigración disminuyen en el año 2020 debido a las restricciones de movilidad por el Covid-19. Por esta razón decidimos no considerar las tasas de emigración de 2020.

Con todo ello, sería buena idea estimar las tasas de emigración de cada provincia como la media de estos años, pero no sabemos si la muestra es homogénea o no. Para comprobarlo, calculamos la media y la mediana correspondiente a todas las provincias y vemos que no coinciden en la mayoría de los casos. Por ello, decidimos estimar las tasas de emigración con la mediana de las tasas de los años 2012 al 2019.

Una vez tenemos todos los datos, multiplicamos las probabilidades por el número total de personas que se marchan de la provincia de origen para determinar el número medio de personas que se desplazan a la provincia destino, es decir, estimamos los flujos migratorios entre provincias. Consideramos por ejemplo las provincias de Albacete y Lugo, de la que tenemos los siguientes datos:

Provincia	Población	Tasa de emigración	Número de emigrantes ( $T_i$ )
Albacete	386.464	12,89	4.982
Lugo	326.013	10,25	3.342

Conociendo el número total de emigrantes de cada una estimamos los flujos migratorios obteniendo:

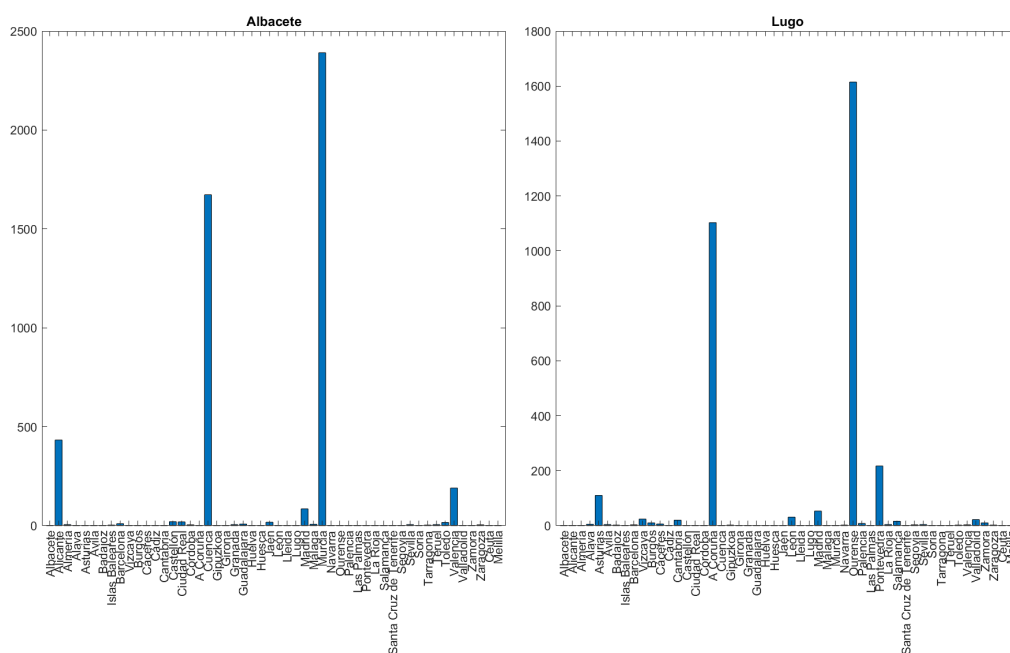


Figura 4: Flujos migratorios con origen la provincia de Albacete y Lugo respectivamente.

Observamos que gran parte de los emigrantes de ambas se trasladan a las provincias colindantes y en menor medida a aquellas que, aunque estén algo más lejos, su población es mayor. Esto verifica nuevamente la teoría.

Con estos valores podemos obtener además el número medio total de migrantes que se va de una provincia origen al resto de provincias y también podemos conocer el número medio total de migrantes que llegan a esta misma provincia desde el resto. Así, el saldo migratorio medio para la migración interprovincial será la diferencia de estos dos valores obteniendo:

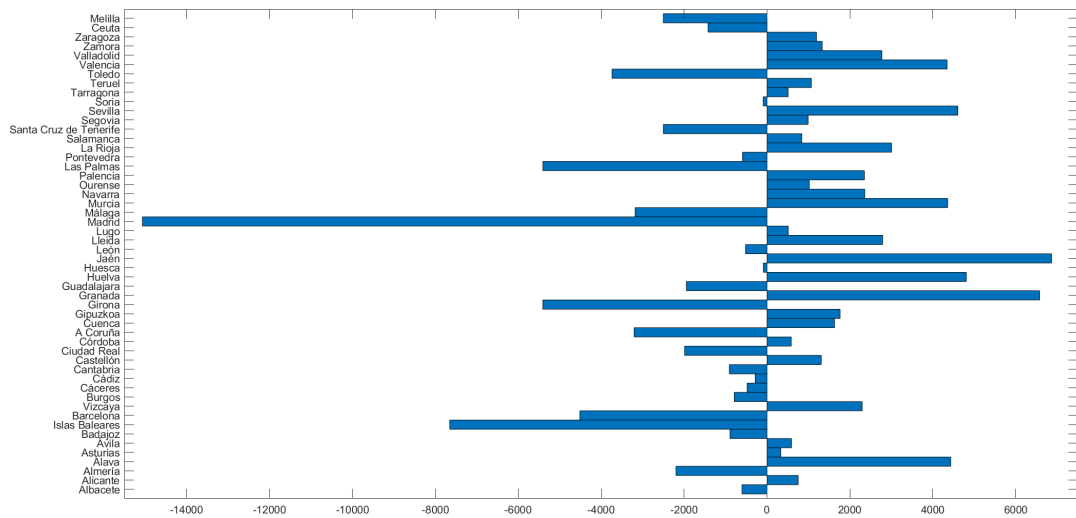


Figura 5: Saldo migratorio medio para la migración interprovincial.

Para calcular los flujos migratorios hemos creado la siguiente función en Matlab:

```
function F = flujosmigratorios(T)

% DATOS DE ENTRADA:
% T es una tabla con 5 columnas en la que se almacenan los
% datos que necesitamos para calcular los flujos migratorios.
% Las columnas reciben el nombre de Provincia, Poblacion,
% Latitud, Longitud y Tasaemigacion respectivamente.

% DATOS DE SALIDA:
% F es la matriz con los flujos migratorios entre regiones

% Calculamos la matriz de distancias utilizando la funcion
% FHaversine
D = zeros(52,52);
for i = 1:52
    for j = 1:52
        if i == j
            D(i,j) = 0;
        else
            D(i,j) = FHaversine(T.Latitud(i),T.Longitud(i),T.
                Latitud(j),T.Longitud(j),6371);
        end
    end
end
end
```

```

% Calculamos la matriz cuyos elementos son la poblacion sij en el
% circulo de centro la region i y radio la distancia entre las
% regiones i y j utilizando la funcion poblacioncirculo
S = zeros(52,52);
for i = 1:52
    for j = 1:52
        if i == j
            S(i,j) = 0;
        else
            S(i,j) = poblacioncirculo(i,j, D, T.Poblacion);
        end
    end
end

% Calculamos la matriz cuyos elementos son la probabilidad pij
% de que una persona se desplace de la region i a la region j
% utilizando la funcion probabilidad
P = zeros(52,52);
for i = 1:52
    for j = 1:52
        if i ~= j
            P(i,j) = probabilidad(i,j,T.Poblacion,S);
        end
    end
end

% Calculamos el vector del numero total de emigrantes de cada
% region
Ti = zeros(1,52);
for i = 1:52
    Ti(i) = round((T.Poblacion(i)*T.TasaEmigracion(i))/1000);
end

% Calculamos la matriz cuyos elementos son el numero medio de
% personas Tij que se desplazan de la region i a la region j
F = zeros(52,52);
for i = 1:52
    for j = 1:52
        if i ~= j
            F(i,j) = floor(Ti(i)*P(i,j));
        end
    end
end
end

```

Con el fin de evitar que la función sea muy larga, creamos una serie de funciones que se utilizan dentro de ella:

```

% Funcion que pasa de grados a radianes
function rad=radianes(grados)
    rad=grados*(pi/180);
end
function dist=FHaversine(lat1,lon1,lat2,lon2,r)

    % DATOS DE ENTRADA:
    % lat1 y lon1 son la latitud y longitud de la region de origen
    % lat2 y lon2 son la latitud y longitud de la region de
    % destino
    % r es el radio de la Tierra

    % DATOS DE SALIDA:
    % dist es la distancia en km entre dos las dos regiones
    % calculada mediante la formula de Haversine

    % Pasamos de grados a radianes utilizando la funcion radianes
    lat=radianes(lat2-lat1);
    lon=radianes(lon2-lon1);
    lat1=radianes(lat1);
    lat2=radianes(lat2);
    a=(sin(lat/2)^2)+cos(lat1)*cos(lat2)*(sin(lon/2)^2);
    c=2*asin(sqrt(a));
    dist=c*r;
end

function sij=poblacioncirculo(posorigen,posdestino, D, col)

    % DATOS DE ENTRADA:
    % posorigen es el numero de fila correspondiente a la region
    % de origen
    % posdestino es el numero de fila correspondiente a la region
    % de destino
    % D es la matriz cuyos elementos son las distancias entre
    % regiones
    % col es la columna Poblacion de la tabla de datos

    % DATOS DE SALIDA:
    % sij es la poblacion total en el circulo de centro la
    % region i y radio la distancia entre las regiones i y j

    sij=0;
    for i=1:52
        if posorigen ~= i && posdestino ~= i && D(posorigen,i)<=D(
            posorigen,posdestino)
            sij=sij+col(i);
        end
    end
end
end

```

```

function pij=probabilidad(posorigen, posdestino, col, S)

% DATOS DE ENTRADA:
% posorigen es el numero de fila correspondiente a la
% region de origen
% posdestino es el numero de fila correspondiente a la region
% de destino
% col es la columna Poblacion de la tabla de datos
% S es la matriz cuyos elementos son las poblaciones sij que
% hay en los circulos de centro la region i y radio la
% distancia entre las regiones i y j

% DATOS DE SALIDA:
% pij es la probabilidad de que una persona que emigra de la
% region i se desplace a la region j

pij=(col(posorigen)*col(posdestino))/((col(posorigen)+S(
    posorigen, posdestino))*(col(posorigen)+S(posorigen,
    posdestino)+col(posdestino)));
end

```

Adicionalmente creamos otra función que, introduciendo los nombres de las provincias de origen y de destino junto con la matriz de flujos migratorios, nos devuelva el flujo migratorio correspondiente:

```

function Tij = flujotij(origen, destino, F, col)

% DATOS DE ENTRADA:
% origen es el nombre de la region de origen (string)
% destino es el nombre de la region de destino (string)
% F matriz de flujos migratorios
% col es la columna Provincia de la tabla de datos

% DATOS DE SALIDA:
% Tij es el flujo migratorio entre las regiones de origen y
% destino

% Buscamos las filas correspondientes a las regiones de origen
% y destino
posorigen = 0;
posdestino = 0;
for i=1:52
    if strcmp(col(i), origen)
        posorigen = i;
    elseif strcmp(col(i), destino)
        posdestino = i;
    end
    if posorigen ~= 0 && posdestino ~= 0
        break
    end
end

% Buscamos el elemento F(posorigen, posdestino)
Tij = F(posorigen, posdestino);
end

```

Como hemos visto en teoría, el modelo de radiación no considera otros factores que no sean económicos y que también puedan influir en la decisión de migrar. Por ello, se propuso el modelo de radiación generalizado. A continuación, realizaremos la misma predicción que antes pero a través del modelo generalizado considerando 4 factores:

1. Densidad de población obtenida dividiendo la población total de cada provincia en 2021 tomada del INE ([10]) entre la superficie de la provincia correspondiente. Estas superficies las tomamos del Instituto Geográfico Nacional (IGN [17])
2. La tasa de paro de cada provincia obtenida del INE ([18])
3. El salario medio de cada provincia, considerando el mismo salario para todas las provincias de la misma Comunidad Autónoma. Estos datos los obtenemos también del INE ([19]), excepto el salario medio de Ceuta y Melilla que los obtenemos de [20]
4. El índice global de Calidad de vida, considerando la misma para todas las provincias de la misma Comunidad Autónoma, obtenidas también del INE ([21]). Este valor lo obtienen como la media del valor dimensión de las condiciones materiales de vida, el trabajo, la salud, la educación, el ocio, la seguridad, la gobernanza y los derechos básicos, el entorno y medioambiente y la experiencia general de vida.

Lo primero que tenemos que hacer es normalizar todos los datos. Debido a la gran diferencia entre algunos de estos, los normalizamos calculando primero el  $\ln(1+x)$  y después aplicamos los tres métodos vistos en la sección anterior para después seleccionar cual funciona mejor.

De los tres métodos, descartamos directamente el método de puntuación  $z$  pues se obtienen valores mayores que 1 en algunas provincias. Comparando los otros dos métodos, observamos que en el método Min-max los valores atípicos toman el valor 0 o 1 mientras que si utilizamos la función logística obtenemos valores en el intervalo  $(0,1)$ . Para que en todas las provincias afecten, en mayor o menor medida, todos los factores optamos por utilizar los valores obtenidos al emplear la función logística para calcular el índice de urbanización de cada provincia.

Para calcular los índices de urbanización necesitamos los pesos. Es lógico que las provincias con mayor tasa de paro no impulsen a la población a migrar a ella, por lo que el valor del peso de ese factor será negativo. Suponemos que es  $w_2 = -10$ . El resto de pesos correspondientes a los tres factores restantes serán positivos. Suponemos que la población busca más una mejor calidad de vida que un salario medio o una densidad de población mayor, por ejemplo, tomamos  $w_1 = w_3 = 25$  y  $w_4 = 50$ .

Conociendo la normalización de los datos correspondientes a los cuatro factores y los pesos calculamos el índice de urbanización de cada provincia obteniendo:

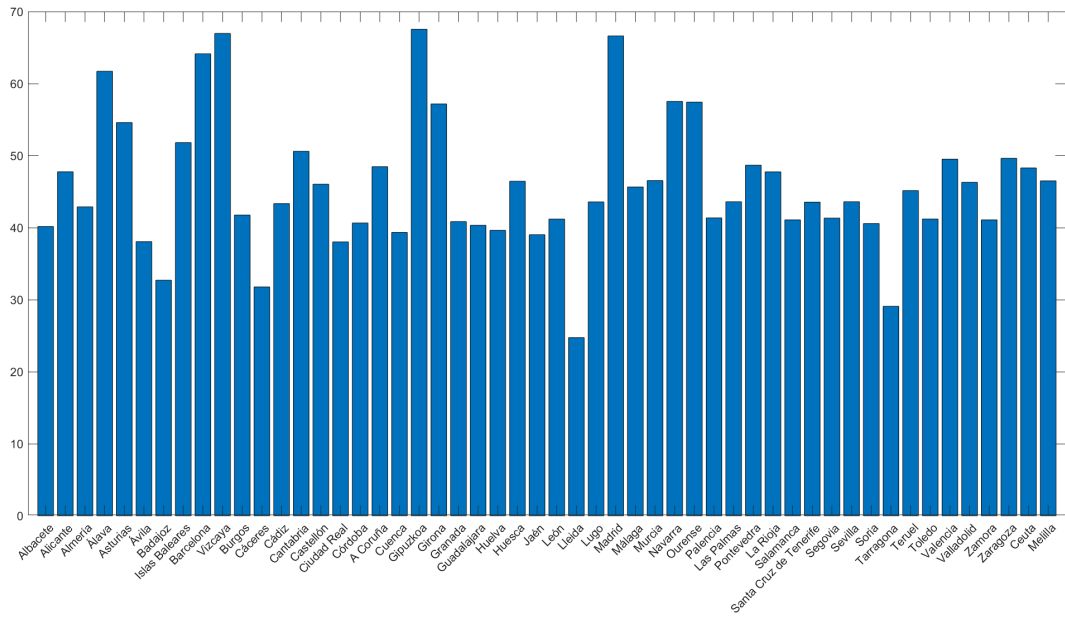


Figura 6: Índices de urbanización de las 52 provincias españolas utilizando cuatro factores:  $f_1$  densidad de población,  $f_2$  tasa de paro,  $f_3$  salario medio,  $f_4$  índice global de calidad de vida con pesos,  $w_1 = 25$ ,  $w_2 = -10$ ,  $w_3 = 25$  y  $w_4 = 50$  respectivamente.

En este caso no empleamos la población total de cada provincia sino el índice de urbanización de estas, por lo que a la función del modelo de radiación no la podemos aplicar tal y como la definimos antes. Para poder utilizarla tenemos que hacer tres pequeñas modificaciones: cambiamos la primera columna de la tabla por el índice de urbanización, introducimos en la tabla una sexta columna con el número de emigrantes de cada provincia y eliminamos la parte en la que lo calculábamos con la tasa de emigración dentro de la función.

Una vez realizadas estas modificaciones, calculamos los flujos migratorios entre las provincias, considerando el índice de urbanización de la figura 6. Tomamos al igual que en el modelo de radiación las provincias de Albacete y Lugo de las que tenemos los siguientes datos:

Provincia	Índice de urbanización	Número de emigrantes ( $T_i$ )
Albacete	40,164	4.982
Lugo	43,575	3.342



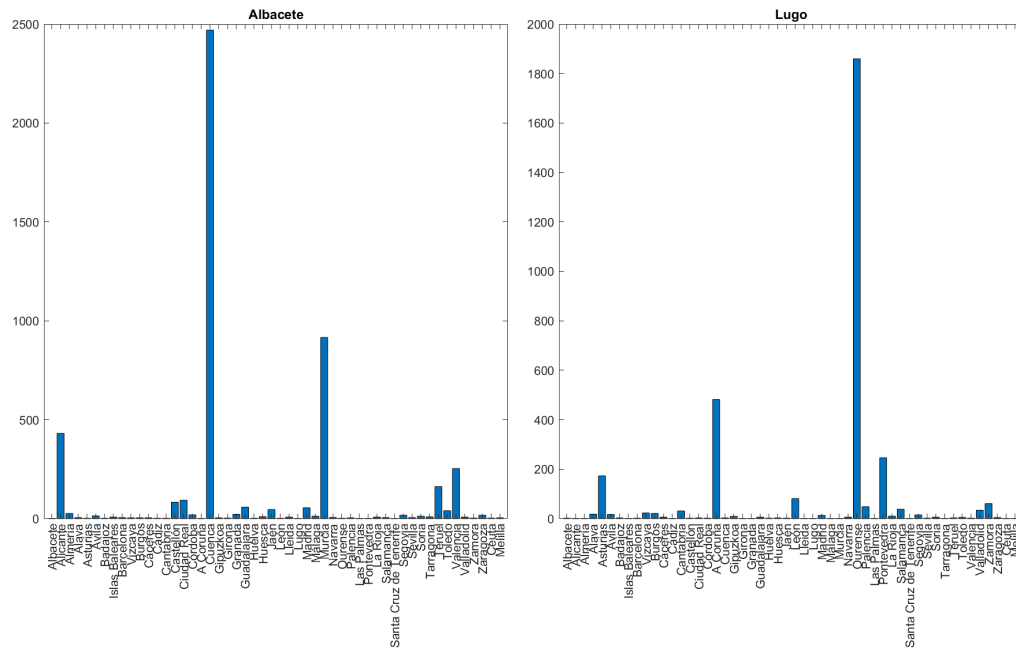


Figura 8: Flujos migratorios con origen la provincia de Albacete y Lugo respectivamente, con pesos de los cuatro factores  $w_1 = 25$ ,  $w_2 = -10$ ,  $w_3 = 50$  y  $w_4 = 50$ .

Observamos que, en el caso de Albacete, hay una diferencia considerable entre los flujos migratorios con las provincias de Cuenca y Murcia, mientras que los flujos migratorios entre la provincia de Lugo y el resto de las provincias se mantienen más o menos con los mismos valores.

Por último, vemos de la misma forma que en el modelo de radiación el saldo migratorio medio para la migración interprovincial de ambos casos y lo representamos en las figuras 9 y 10. Observamos que, como hemos visto antes, al modificar los pesos de los factores también se modifican, en mayor o menor medida, los flujos migratorios entre provincias afectando lógicamente también al saldo migratorio. Además, entre ambos se observa que el signo del saldo migratorio se conserva en las provincias mientras que si lo comprobamos con el saldo migratorio medio del modelo de radiación, en alguna provincias como Ceuta el signo es opuesto.

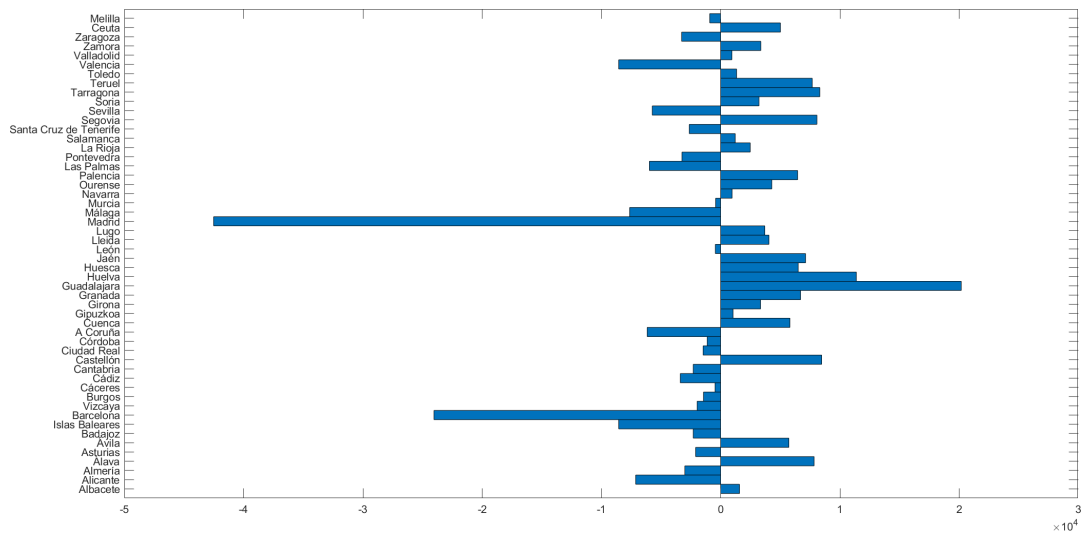


Figura 9: Saldo migratorio medio para la migración interprovincial considerando los cuatro factores con pesos  $w_1 = 25$ ,  $w_2 = -10$ ,  $w_3 = 25$  y  $w_4 = 50$  respectivamente.

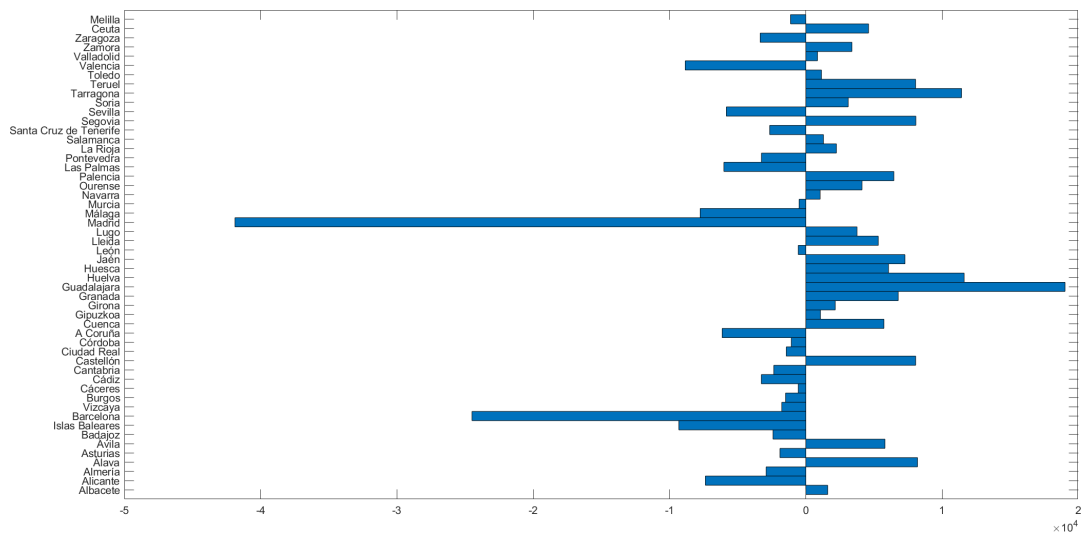


Figura 10: Saldo migratorio medio para la migración interprovincial considerando los cuatro factores con pesos  $w_1 = 25$ ,  $w_2 = -10$ ,  $w_3 = 50$  y  $w_4 = 50$  respectivamente.

## 5 Conclusión

A lo largo del trabajo hemos visto algunos de los muchos modelos propuestos para el estudio de los fenómenos migratorios, clasificándolos en función de la teoría en la que se basan para desarrollarlo. Entre los modelos de un mismo bloque existen similitudes, pero también diferencias debido a que la mayoría se proponen como mejora de otro anterior.

A pesar de la gran variedad de modelos, ninguno modeliza de forma cien por cien fiable los fenómenos migratorios. Esto no tiene por qué deberse únicamente al propio modelo, sino que también puede verse influido por la escasez y dificultad para conocer los datos. Además, la decisión de migrar a otra región también se debe a factores que no pueden ser medidos o que son difíciles de estimar, lo que aumenta la complejidad de su estudio.

Todo ello lo podemos observar comparando los resultados de la simulación realizada en la sección anterior. En ella, primero calculamos los flujos migratorios entre las provincias españolas utilizando el modelo de radiación. Para ello, conocemos la población total de cada una de ellas junto con sus tasas de emigración. Después obtuvimos los flujos migratorios entre las provincias empleando el modelo de radiación generalizado para dos casos, considerando en ambos los mismos factores pero con distintos pesos.

Si observamos las figuras 5, 9 y 10, vemos que entre el modelo de radiación y el modelo de radiación generalizado hay diferencias en el saldo migratorio medio y por lo tanto en los flujos migratorios. En principio, podemos pensar que es debido a que consideramos distintos factores en ambos modelos. Sin embargo, si comparamos la figura 9 y 10 también es posible encontrar diferencias en los flujos migratorios, aunque en menor medida.

Esto corrobora que en función de los datos que tengamos y de los factores que consideremos obtendremos estimaciones diferentes de los mismos flujos migratorios. La gran cantidad de factores que pueden influir en los movimientos migratorios hace difícil tener una claridad conceptual sobre este fenómeno, por lo que es muy complejo su estudio matemático.

## 6 Bibliografía

- [1] Cross, J. G. (1978). *A stochastic learning model of migration*. Journal of Development Economics, 5(2), 195-202.
- [2] Martínez, Ú. F., García, J. G. (2001). *Modelos migratorios: Una revisión*. RAE: Revista Asturiana de Economía, (21), 209-235.
- [3] Martínez, U. F., García, J. G., Gallego, J. A. (2000). *Estudio de la migración interregional en España, a través de la Ecuación Master*. Estudios de Economía Aplicada, 16(3), 63-92.
- [4] Franco, M. H., Garvi, M. G. (2007). *Una revisión de la aplicación de las cadenas de Markov discretas al estudio de la movilidad geográfica*. Estadística española, 49(166), 473-499.
- [5] Santos Preciado, J. M. (1994). *Los modelos de interacción espacial y el análisis de los flujos migratorios interregionales: aplicación al territorio español*.
- [6] Todaro, M. P. (1969). *A model of labor migration and urban unemployment in less developed countries*. The American economic review, 59(1), 138-148.
- [7] Simini, F., González, M. C., Maritan, A., Barabási, A. L. (2012). *A universal model for mobility and migration patterns*. Nature, 484(7392), 96-100.
- [8] de Vries, J. J., Nijkamp, P., Rietveld, P. (2001). *Alonso's theory of movements: developments in spatial interaction modeling*. Journal of Geographical Systems, 3(3), 233-256.
- [9] Gamboa Pérez, M. (2017). *Estudio en tiempo discreto de la expansión de una epidemia*.
- [10] <https://www.ine.es/jaxiT3/Datos.htm?t=2852>
- [11] Rivera, F., Angel, L. (2012). *Teorema de Perron-Frobenius Aplicado a las Matrices Estocásticas* (Doctoral dissertation).
- [12] <https://www.ine.es/jaxiT3/Datos.htm?t=3996>
- [13] Chopde, N. R., Nichat, M. (2013). *Landmark based shortest path detection by using A\* and Haversine formula*. International Journal of Innovative Research in Computer and Communication Engineering, 1(2), 298-302.
- [14] <https://www.ine.es/jaxiT3/Datos.htm?t=5956>
- [15] Alis, C., Legara, E. F., Monterola, C. (2021). *Generalized radiation model for human migration*. Scientific reports, 11(1), 1-10.
- [16] González, W. J. G., Flórez, J. J. M., Londoño, S. M. P. (2014). *Análisis del procesamiento de los datos de entrada para un localizador de fallas en sistemas de distribución*. Tecnura, 18(41), 64-75.

- [17] <https://web.archive.org/web/20220208134600/https://www.ign.es/web/ign/portal/ane-datos-geograficos/-/datos-geograficos/datosPoblacion?tipoBusqueda=provincias>
- [18] <https://www.ine.es/jaxiT3/Datos.htm?t=49170>
- [19] <https://www.ine.es/jaxiT3/Datos.htm?t=28191>
- [20] <https://datosmacro.expansion.com/ccaa/comparar/ceuta/melilla?sc=XE15>
- [21] [https://public.tableau.com/views/IMCV\\_caaES/Dashboard2?showVizHome=no:embed=true3](https://public.tableau.com/views/IMCV_caaES/Dashboard2?showVizHome=no:embed=true3)