

# Proyección de la población mexicana con Series de Tiempo Difusas: horizonte 2100

## Projection of the Mexican population with Fuzzy Time Series: horizon 2100

Milenka Linneth Argote-Cusi y León Darío Parra-Bernal

*Business Intelligence and Demography, Colombia  
Universidad EAN, Colombia*

### *Resumen*

La proyección de la población es crucial para la planificación de recursos en todos los países. A diferencia de los métodos tradicionales, las series de tiempo difusas (STD) es una técnica de la IA que aprende de los datos, los modela incorporando la lógica difusa y proyecta considerando el comportamiento no lineal del fenómeno. Objetivo. Actualizar la proyección de la población mexicana con STD y realizar un análisis comparativo de resultados de 2018 *versus* 2022 así como con otras estimaciones disponibles en la literatura. Datos y Métodos. Se considera la serie de datos oficiales de la población total mexicana de 1895 a 2020 y los resultados de los autores en 2018 que se toman como referencia para el análisis de la precisión de las estimaciones; se generó un algoritmo en R que ejecuta los pasos de la metodología de STD invariantes en el tiempo. Resultados. Se comprueba que las STD permiten modelar de forma adecuada la dinámica de la población mexicana en la que los resultados son coherentes. El ejercicio de proyección en la etapa retrospectiva (1950 a 2020) tiene un menor error 0,03409 que en 2018 (0,04074) y la inclusión del dato de 2020 nos permite ampliar el horizonte de proyección a 2100. La validez y veracidad de los resultados se confirman con el análisis comparativo con otras proyecciones de la población mexicana.

*Palabras clave:* Series de Tiempo Difusas, proyecciones de población, México, lógica difusa, precisión.

### *Abstract*

Population projection is crucial for resource planning in all countries. Unlike traditional methods, Fuzzy Time Series (FTS) is an AI technique that learns from the data, models it incorporating fuzzy logic, and projects considering the nonlinear behavior of the phenomenon. Goal. Update the projection of the Mexican population with STDs and perform a comparative analysis of the results of 2018 versus 2022 as well as with other estimates available in the literature. Data and Methods. The official data series of the total Mexican population from 1895 to 2020 and the results of Authors in 2018 are considered, which are taken as a reference for the analysis of the precision of the estimates; an algorithm was generated in R that executes the steps of the time-invariant STD. Results. It is verified that the STD allow to adequately model the dynamics of the Mexican population, the results are consistent. The projection exercise in the retrospective stage (1950 to 2020) has a lower error of 0.03409 than in 2018 (0.04074) and the inclusion of the 2020 data allows us to extend the projection horizon to 2100. The validity and veracity of the results are confirmed with the comparative analysis with other projections of the Mexican population.

*Key words:* Fuzzy Time Series, population forecasting, Mexico, fuzzy logic, accuracy

## INTRODUCCIÓN

**L**a proyección de datos o el pronóstico son de interés sustancial de todas las disciplinas para conocer el futuro incierto. A pesar de que es difícil predecir, y que los modelos de pronóstico del clima han demostrado poca certeza, más aún hoy con el cambio climático, el ejercicio de modelar diferentes fenómenos y generar estimaciones reducen la incertidumbre y son información valiosa para la toma de decisiones y la planificación de recursos (Alho, Alders, Cruijssen, Keilman, Nikander y Pham, 2006; Argote y Parra, 2020). El mejor ejemplo de la utilidad de los modelos de proyección en la actualidad son los ejercicios de proyección del Covid-19 generados por el Institute of Health Metrics and Evaluation (IHME) de la Universidad de Washington que se han tomado como referencia a nivel mundial para la toma de decisiones con relación al comportamiento de la pandemia y que ha permitido planificar los recursos en salud para que el sistema sanitario no colapse.

Existen varios métodos de proyección de datos. Desde el modelo más sencillo de regresión lineal que ajusta una recta a datos que tienen comportamiento lineal, hasta el uso de técnicas más sofisticadas como los modelos ARIMA que consideran ecuaciones más complejas. El reto de la proyección de datos es modelar comportamientos complejos similares a la realidad, lo cual divide a los métodos en tradicionales y no tradicionales. Los primeros, centrados en la linealidad y los supuestos como el equilibrio y la constante que fueron el punto de partida de metodologías que buscaban la sencillez y la practicidad a la hora de proyectar en desmedro de capturar una dinámica más compleja. El miedo a modelar sistemas complejos se fue perdiendo a medida que los modelos teóricos que iniciaron primero, se pudieron implementar de forma empírica gracias a los avances tecnológicos y la computación. Con el pasar de los años, la perspectiva de los modelos estocásticos, los modelos multiagente y la inteligencia artificial para el modelado de fenómenos sociales se ha hecho frecuente. Indudablemente este salto ha requerido la integración de técnicas provenientes de la matemática, ingeniería y estadística que junto con herramientas computacionales, hoy en día han hecho realidad la implementación de estos modelos (Aladag, Aladag, Mentés y Egrioglu, 2012; Sullivan y Woodall, 1994; Argote y Parra, 2020).

En este contexto científico sobre los métodos de proyección, el principal aporte de la presente investigación es actualizar la proyección de la población mexicana utilizando Series de Tiempo Difusas (STD) que rea-

lizó Argote en 2016 y se publicó en 2018, incluyendo el dato de 2020 a la serie para evaluar la precisión de las estimaciones siete años después, con una versión automatizada del algoritmo de predicción desarrollado en R, lo cual permite realizar varias simulaciones de tal forma de alcanzar los resultados que más se ajusten a los datos observados. Ello permite contar con un ejercicio de proyección a 2022 para evaluar las diferentes estimaciones con métodos tradicionales encontradas en la literatura y con STD realizadas desde 2016 para el caso de México.

El resto del artículo está organizado de la siguiente forma. En la sección 2 se presenta una breve inmersión a los métodos de proyección en general y en la sección 3 se profundiza en la trayectoria de las Series de Tiempo Difusas (STD) para proyectar datos. En la sección 4 se presentan los datos utilizados en la investigación. En la sección 5, se presenta el método de proyección y en la sección 6 el algoritmo implementado R con las funciones que lo componen. En la sección 7 se presentan los resultados de la proyección con STD y otras estimaciones y finalmente en la sección 8 las conclusiones.

## LA CIENCIA DETRÁS DE LA PROYECCIÓN DE DATOS

Los métodos de proyección son fundamentales para la planificación y la toma de decisiones bajo incertidumbre. La pandemia por Covid-19, que inició en 2020, nos ha demostrado que cualquier evento por poco probable que parezca puede ocurrir. Ante esta incertidumbre cobran relevancia los métodos de proyección, ya que, si bien no podemos predecir totalmente el futuro, si podemos reducir la incertidumbre generando escenarios probables para tomar decisiones más acertadas.

Los métodos de proyección forman parte del grupo de métodos matemáticos sofisticados en varias áreas. Para la ciencia de los datos, la perspectiva es la capacidad de más alto nivel en la escala de madurez de las organizaciones con relación al uso de los datos para la toma de decisiones (Kopsco y Pachamanova, 2018). En el ámbito estadístico, los métodos de proyección se ubican en el nivel de inferencia más alto. En demografía, la proyección de la población se encuentra en el top de la pirámide de los métodos actuariales que modelan el comportamiento de la población, con el objetivo de contar con evidencia científica para la planificación y asignación de los recursos escasos a nivel local y nacional (Argote, 2018; Raftery y Ševčíková, 2021; Argote y Parra, 2020).

Desde la ciencia estadística, la regresión lineal simple es el método más sencillo de proyección, el cual ajusta una recta a un conjunto de datos

que se asume tenga un comportamiento lineal, es decir, que represente una relación lineal creciente o decreciente entre dos variables. Por otro lado, el “método por componentes” en demografía considera la estructura de edades, las tasas de fecundidad, mortalidad y de migración para proyectar la población (CEPAL, 2009). Se trata de un método en esencia matricial en el cual los niveles de población son afectados por tasas de fecundidad, mortalidad y migración que son las variables más importantes que afectan el nivel de la población (Raftery y Ševčíková, 2021).

En otro ámbito, tenemos los métodos de proyección desde la perspectiva de la econometría. En este grupo tenemos los modelos ARMA, ARIMA y SARIMA, cuyo objetivo es modelar series de tiempo univariantes estacionales o no estacionales. Adicionalmente, el modelo de Box y Jenkins, incorporan una variable subjetiva que permite ser adaptable y elegir entre varios modelos (Aqil and Akhter, 2005). Una de las desventajas de estos modelos complejos es precisamente su complejidad, ya que al incorporar muchas variables y parámetros es muy probable que crezca el error atado a ellos, así como se pueden presentar problemas de heterocedasticidad y endogeneidad por lo que requieren diferentes tipos de test y ajustes. En la literatura se encuentra su utilidad en diversos casos incluso comparado con otros métodos (Wang, Li y Lim, 2019).

En la era digital, la evolución de la computación ha dado origen al desarrollo de nuevos métodos como la Inteligencia Artificial (IA). La IA desde su definición en 1956, por Martin Minsky y John McCarthy, ha evolucionado hasta convertirse hoy en día en una ciencia que estudia teorías, métodos y técnicas para simular la inteligencia humana. La inteligencia artificial, a diferencia de los métodos tradicionales, hace uso de las capacidades de cómputo para modelar y proyectar datos basado en modelos neuronales. Es así que hoy se disponen de varias técnicas como las redes neuronales, SVM, Fuzzy Logic, etcétera, que han demostrado una mayor precisión de los resultados y por ende, una reducción del error atado a las estimaciones (Mustapha, Haruna y Muhammad, 2020). Un ejemplo de ello, es el trabajo de ArunKumar, Kalaga, Kumar, Kawaji y Brenza (2022) quienes realizan la proyección de datos del Covid-19, utilizando modelos estadísticos (ARIMA y SARIMA) y modelos de redes neuronales (GRU y LSTM), lo cual resulta en un análisis comparado muy interesante que evidencia, según las medidas de error, que los modelos de *Deep Learning* (DL) tienen mejores desempeños.

Los principales desarrollos de la IA para la proyección de datos se han dado de forma acelerada entre 2013 y 2019 en el área económica (Nosratabadi, Mosavi, Duan, Ghamisi, Filip, Band, Reuter, Gama y Gandomi, 2020). Así tenemos a Souma, Vodenska y Aoyama (2019) que utilizan técnicas de IA para mejorar la proyección de datos financieros del mercado a través del análisis de sentimientos. En 2022, Li y Pan proponen un conjunto de modelos de *Deep Learning* para la predicción de las acciones en el mercado basado en los precios y las noticias. Bajo el conocimiento que el precio de las acciones es una variable que posee muchos factores de influencia. Los investigadores proponen acercarse a factores poco estudiados como las noticias y las series de tiempo, para proyectar el comportamiento de las acciones con mayor precisión (Nosratabadi, Mosavi, Duan, Ghamisi, Filip, Band y Gandomi, 2020; Pan, Yang, Zhou y Kong, 2020).

Sin embargo, la proyección de datos económicos es muy diferente a los datos poblacionales. Mientras las series de datos económicos como el desempleo y el Producto Interno Bruto (PIB) requieren menor tiempo de estimación, la dinámica de las poblaciones posee una inercia diferente, lo que permite que se pueda proyectar a mayores horizontes en el futuro como lo explican al detalle Raftery y Ševčíková (2021). Es de esta manera que la proyección de datos poblacionales ha evolucionado en el tiempo desde el método clásico “por componentes” (CCMPP por sus siglas en inglés *cohort-component method of population projection*) que se ha utilizado de forma estándar desde 1940. Es solo a partir de la última década en que se plantean modelos probabilísticos ante la necesidad de incorporar el error y la incertidumbre en las estimaciones de población (Raftery y Ševčíková, 2021; Ordorica, 2004).

En este contexto, la presente investigación se enmarca en los métodos de proyección con Series de Tiempo Difusas (STD) para la proyección de datos poblacionales de la cual se tiene preferencia desde 2003 con el trabajo de Abbasov y Mamedova y el trabajo de Argote en 2018 para la proyección de la población mexicana. A diferencia de los trabajos exhaustivos en mejorar la precisión de los intervalos y de reducir el proceso de cómputo de las operaciones matriciales del método, la presente investigación crea su propio algoritmo para simular el método con diferentes datos, lo cual permite realizar un análisis comparativo exhaustivo de la precisión de las estimaciones de la población mexicana a 2100, tomando en cuenta los hallazgos de Argote en 2018.

## USO DE SERIES DE TIEMPO DIFUSAS PARA LA PROYECCIÓN DE FENÓMENOS SOCIALES

La lógica difusa fue desarrollada por Lotfi A. Zadeh en 1965 en su interés de modelar el lenguaje natural y sus matices, es decir, más allá de ver al mundo en blanco y negro, la lógica difusa toma en cuenta una gama de colores, matices de grises entre el 0 y el 1, lo cual le brinda al modelo una mayor representatividad de la realidad (Zadeh, 1973; Mustapha, Haruna y Muhammad, 2020). En 1993 Song y Chissom desarrollaron a partir de los conceptos teóricos sobre lógica difusa la formulación matemática de las series de tiempo difusas como método de proyección.

La ventaja de las series de tiempo difusas con relación a otros métodos es su sencillez y su acertada representación de los matices de la realidad al incorporar la no linealidad de la función de membresía. A partir de una serie de datos, es capaz de aprender de la serie histórica pasada, entrenar el modelo bajo una función de membresía, realizar ajustes y utilizar el mismo para proyectar el futuro. De forma básica, el método consiste en seleccionar una serie de datos del pasado, encontrar las (1) variaciones entre pares de datos en el tiempo  $t$  y el anterior  $t - 1$ , con base en las variaciones encontrar el rango en que fluctúan lo cual constituye el universo de discurso, se define un (2) conjunto de intervalos de referencia que, con base a una (3) función de membresía asigna un valor a cada dato de la serie lo cual se llama fuzzificación. Finalmente, se realiza un proceso de (4) defuzzificación que transforma el valor fuzzificado en un valor coherente a la serie de datos. Si el modelo brinda buenos niveles de precisión con la serie histórica, se procede a generar la serie futura.

A lo largo del tiempo, la semilla originada en 1993 ha tenido muchos aportes en los diferentes pasos del método de proyección. El primero en buscar una mejora al método de Song y Chissom fue Chen (1996) para reducir las capacidades de cómputo que demandaban las operaciones entre matrices (max-min); plantea un método heurístico a través de operaciones aritméticas sucesivas con las cuales logra mejorar las estimaciones utilizando la serie de datos del número de matriculados de la Universidad de Alabama. Song y Chissom (1993 y 1994) mejoran sus modelos anteriores incluyendo series de tiempo variantes en el tiempo y redes neuronales.

Posteriormente varios autores se centran en la división de los intervalos con diferentes metodologías. Huarng (2001) propone la mejora en la definición del tamaño de los intervalos a través de un promedio, mientras que Chen y Hsu (2004) dividen de forma uniforme el universo de discurso para

obtener la distribución estadística en cada intervalo y dividir nuevamente, finalmente Li y Chen (2004) aplican una técnica de partición natural nivel por nivel. Burney, Ali and Khan (2018) en su propuesta incorporan el método de Huarng para la división de los intervalos del universo de discurso, una función trapezoidal para las membresías y plantean un modelo de segundo orden para la serie de datos difusos.

Desde el punto de vista de la aplicación, se encuentran antecedentes del uso de series de tiempo difusas para la predicción de la temperatura (Chen y Hwang, 2000). En 2003 Abbasov y Mamedova utilizan series de tiempo difusas invariantes para la proyección de la población de Azerbaiyán de 2002 a 2012 basada en los datos históricos entre 1988 y 2001, obteniendo errores que fluctúan entre 0.02 y 0.25 con un promedio de 0.13. Por otro lado, Sasu (2010) aplicó la metodología utilizada para la población de Rumania entre 1988-2009 y encontró resultados satisfactorios; se evidencian errores menores a 0.003 entre los datos observados y proyectados. Del otro lado del continente, Argote (2018) desarrolla su propio algoritmo para proyectar la población mexicana con Fuzzy Time Series (FTS) tomando la serie histórica de 1895 a 2010 para proyectar de 2010 a 2050, encontrando un error promedio de 0.0407 en la proyección de la población; los resultados absolutos se comparan con cinco estimaciones de entidades oficiales del gobierno y muestran resultados coherentes y con menor error.

Tai Vovan (2019), basado en el modelo de Abbasov y Mamedova (2003) hace una nueva propuesta incluyendo en el modelo de STD, algoritmos para la estimación de dos parámetros importantes: la constante C (Algorithm to find the appropriate value for the constant C) y el número de intervalos en que se debe dividir el universo de las variaciones de la serie de datos (Algorithm to find the dividing intervals for the universal set); calibra el modelo con datos de la producción de cereal en la India de 1966 a 2011 utilizados por Ghosh *et al.* (2015) y posteriormente aplica el modelo para predecir los picos de sal en la provincia Ca Mau del sur de Vietnam que es una de las áreas más afectadas por el cambio climático. Finalmente, utiliza el modelo para la predicción de la población total de Vietnam. Los resultados son comparados con anteriores investigaciones encontrando menor error.

Yusuf, Mohammad y Hamisu (2017) aplican una STD para la proyección de temperaturas, a diferencia de otros, plantean una nueva forma de encontrar los intervalos de los conjuntos difusos a través de un modelo matemático que mejora el desempeño de STD. En el ámbito financiero Gorbatiuk, Hryhoruk, Proskurovych, Rizun, Gargasas, Raupelienè y Mun-

jishvili (2021) tienen el interés de predecir el nivel de ingreso neto del sector empresarial en Ucrania entre 2002 y 2017, así como aplican el método de Stevenson y Potter para considerar las variaciones como una tasa de crecimiento.

El interés por la mejora de la precisión del método de proyección en sus diferentes fases: determinación efectiva de los intervalos, fuzificación de la serie de datos, modelado de las relaciones de lógica difusa y defuzificación, ha sido constante (Panigrahi y Behera, 2020). Vovan (2019) y Rana (2020), entre otros, construyen algoritmos cada vez más sofisticados y complejos para mejorar las estimaciones, incluso los combinan con otras técnicas del *soft computing* como las redes neuronales tomando como ejemplo los datos de los matriculados a la Universidad de Alabama entre otros (Bas, Uslu, Aladag, Yolcu y Egrioglu, 2014; Egrioglu, Bas, Aladag y Yolcu, 2016; Ghosh, Chowdhury y Prajneshu, 2015; Chen y Tanuwijaya, 2011).

Como se puede evidenciar, los avances en el uso de la técnica se han realizado en la búsqueda de la precisión, optimizando algunos de los procesos de la metodología. Debido a ello, la presente investigación se centra en aportar en el análisis en profundidad de la metodología para el caso de la población mexicana y sus variaciones a diferentes tipos de simulaciones que se pueden hacer a partir del algoritmo desarrollado.

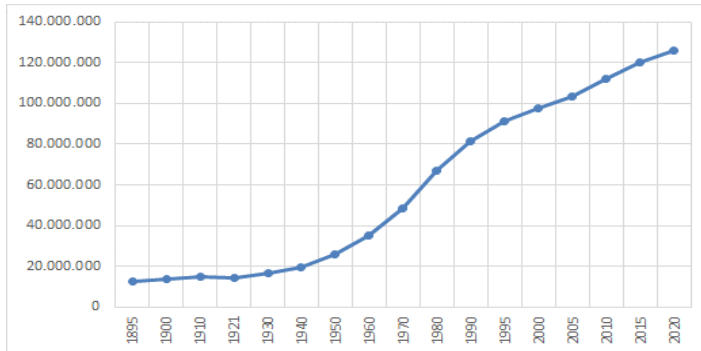
## **DATOS**

México posee una importante historia de generación de información poblacional. Es en este sentido que se dispone de información de acceso libre en las instituciones oficiales de gobierno que centralizan la información poblacional, la cual se utiliza en las diferentes dependencias del estado para la formulación, ejecución y evaluación de políticas públicas del país.

Aprovechando esta fuente rica de información, se obtiene la serie de datos de la población mexicana de 1895 a 2020 del Instituto Nacional de Estadística y Geografía (INEGI) y del Consejo de Población (CONAPO). Cabe resaltar que después de 1990 se planifica cada cinco años un censo (1995, 2005 y 2015) que ha sido de mucha utilidad a la hora de realizar ajustes a las estimaciones, sin embargo, para el modelo STD, por homogeneidad se consideran los datos decenales que corresponden a los censos nacionales.

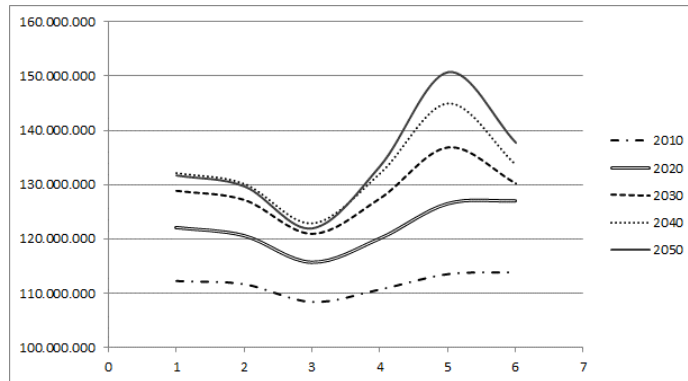


Figura 1: Población de México 1895 a 2020



Fuente: elaboración propia con base datos del INEGI.

Figura 2: Comparación de diferentes estimaciones de la proyección de la población mexicana 2010-2050



Fuente: tomado de Argote (2018).

Por otro lado, para la fase comparativa se utilizan los datos de la investigación de Argote en 2018, que posee resultados de la simulación y proyección de la población mexicana realizada en 2016 (Argote, 2016).

## MÉTODO

Las definiciones y teoremas básicos de las series de tiempo difusas que son esenciales para entender la metodología, se pueden revisar en Argote (2018), sin embargo, a continuación se detalla los pasos:

Sea  $U$  el universo de discurso,  $U = \{u_1, u_2, \dots, u_n\}$  y sea  $A$  en el universo de discurso  $U$  definida de la siguiente manera:

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n \quad (1)$$

Donde  $f_A$  es la función de membresía de  $A$ ,  $f_A \rightarrow [0,1]$ ,  $f_A(u_i)$  indica el grado de membresía de  $u_i$  en el conjunto difuso  $A$ ,  $f_A(u_i) \in [0,1]$  y  $1 \leq i \leq n$ .

Sea  $X(t)$  ( $t = \dots, 0, 1, 2, \dots$ ) el universo de discurso y subconjunto de  $R$ , y sea el conjunto difuso  $f_i(t)$  ( $i = 1, 2, \dots$ ) definido en  $X(t)$ . Sea  $F(t)$  una colección de  $f_i(t)$  ( $i = 1, 2, \dots$ ). Entonces  $F(t)$  es llamada serie de tiempo difusa de  $X(t)$  ( $t = \dots, 0, 1, 2, \dots$ ).

Si  $F(t)$  es causado por  $F(t-1)$ , es decir  $F(t-1) \rightarrow F(t)$ , entonces esta relación puede ser representada por  $F(t) = F(t-1) \circ R(t, t-1)$ , donde el símbolo " $\circ$ " denota el operador compuesto Max-Min;  $R(t, t-1)$  es una relación difusa entre  $F(t)$  y  $F(t-1)$  y es llamado modelo de primer orden de  $F(t)$ .

Sea  $F(t)$  una serie de tiempo difusa y sea  $R(t, t-1)$  el modelo de primer orden de  $F(t)$ . Si  $R(t, t-1) = R(t-1, t-2)$  para cualquier  $t$ , entonces  $F(t)$  es llamada serie de tiempo difusa invariante en el tiempo. Si  $R(t, t-1)$  es dependiente del tiempo, esto es,  $R(t, t-1)$  puede ser diferente de  $R(t-1, t-2)$  para cualquier  $t$ , entonces  $F(t)$  es llamada serie de tiempo difusa variante en el tiempo.

Considerando las definiciones anteriores, el método de proyección se lleva a cabo a través de los siguientes pasos:

Paso 1: Definir el universo de discurso, los conjuntos difusos y las variables lingüísticas.

Paso 2: Particionar el universo de discurso en intervalos iguales.

Paso 3: Determinar los valores de las variables lingüísticas representadas por los intervalos en que se dividió el universo de discurso.

Paso 4: Fuzzificar los datos históricos según la siguiente función de membresía:

$$\mu_{A^{mn}}(u_i) = \frac{1}{1 + [C * (U - u_m^i)]^2} \quad (2)$$

Donde  $A^{mn}$  es el conjunto difuso que corresponde a las variaciones entre los años [2010,2020],  $C$  es una constante que tomará varios valores para su evaluación,  $U$  es el universo de discurso conformado por las variaciones entre la serie de datos y  $u_m^i$  es el punto medio del intervalo  $u_i$ . Los conjuntos difusos se definen en el conjunto universo  $U$ .

Paso 5: Elegir el parámetro  $w$  donde  $w > 1$ , calcular la matriz  $R^w(t, t-1)$  y proyectar la población de la siguiente forma:

$$F(t) = F(t-1) \circ R^w(t, t-1) \quad (3)$$

Donde  $w$  es el dato proyectado de la población en el año  $t$ ,  $F(t-1)$  es la población fuzzificada del año  $t-1$  y

$$R^w(t, t-1) = F^T(t-2) \times F(t-1) \cup F^T(t-3) \times F(t-2) \cup \dots \cup F^T(t-w) \times F(t-w+1) \quad (4)$$

Donde  $w$  es llamado “modelo base” que considera el número de años antes del tiempo  $t$ , “ $\times$ ” es el producto cartesiano y  $T$  es el operador transpuesto.

Paso 6: Defuzificar la población proyectada. Es necesario transformar la respuesta en una forma que no sea difusa, en este caso se utiliza el método del centroide de área (Jang *et al.*, 1997):

$$v(t) = \frac{\sum_{i=1}^5 \mu_t(u_i) \cdot u_m^i}{\sum_{i=1}^5 \mu_t(u_i)} \quad (5)$$

Donde  $\mu_t(u_i)$  es el valor calculado de la función de membresía para el año proyectado  $t$  y  $u_m^i$  es el valor medio del intervalo.

Paso 7: El error estimado de las proyecciones de población acorde a la presente metodología se calculó a través de la siguiente fórmula:

$$\delta(t) = \frac{V_{obs}^t - V_{proy}^t}{N_{obs}^t} * 100 \quad (6)$$

Donde  $V_{obs}^t$  es la variación de la población mexicana en el año  $t$ ;  $V_{proy}^t$  es la variación de la población proyectada en el año  $t$ ;  $N_{obs}^t$  es la población total observada en el año  $t$ ,  $2010 \leq t \leq 2020$ .

### ALGORITMO DE STD-P

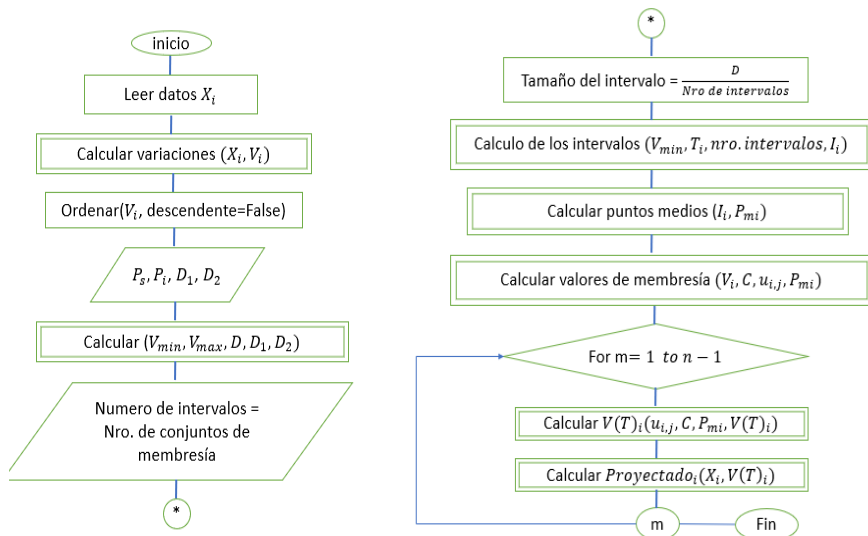
Tomando como referencia la metodología de STD descrita por Song y Chisom (1993) y Abbasov y Mamedova (2003) se crea el algoritmo STD-P considerando varias funciones para proyección de la población mexicana. A continuación, se presenta los pasos del algoritmo:

1. Leer serie de datos:  $x_1, x_2, x_3, \dots, x_n$
2. Numero de datos:  $n$
3. Calcular las variaciones  $V_{i^*}$  entre  $x_1, x_{i+1}$
4. Ordenar variaciones de mayor a menor
5. Definir una cota superior  $P_s$  y una cota inferior  $P_i$  próxima a los límites de universo de discurso.
6. Calcular *distancia* =  $P_s - P_i$ ;  $D_1 = V_{\min} - P_i$ ;  $D_2 = P_s - V_{\max}$

7. Introducir la constante  $C$  de la función de membresía tal que  $C \in \{0, \dots, 1\}$
8. El número de intervalos es igual al número de conjuntos de membresía ( $m$ )
9. Tamaño del intervalo  $distancia/m$
10. Calcular los intervalos
11. Calcular puntos medios de los intervalos  $Pm_{i,*}$
12. Calcular los valores de membresía  $\mu_{i,j}$  con base en  $V_{i,*}$ , los intervalos y (5)
13. Iniciar bucle: recorrer la matriz de valores de membresía
  - a. Calcular  $R(T) = O(T) \circ K(T)$
  - b. Calcular  $F(T) = \max (R(T)_{i,j})$
  - c. Calcular  $V(T)_i$  en función de  $F(T)$  y  $Pm_i$
14. Calcular los valores proyectados en función de la serie

El algoritmo se implementó en el lenguaje de programación R. Para una mayor rapidez de cómputo se diseñaron varias funciones, cada una con un objetivo principal. En la Figura 3 se puede observar el algoritmo “madre” que llamará a las funciones según la necesidad. Se puede observar que se cuenta con cinco funciones. La primera función denominada “variaciones” calcula las variaciones entre los datos de la serie histórica.

Figura 3: Algoritmo principal para la proyección de la serie histórica con STD-P



Fuente: elaboración propia.

En segundo lugar, la función “intervalos”, con base en las variables lingüísticas definidas que se pueden ver en la Tabla 2, se calculan los intervalos con base en las variaciones y sus puntos medios.

Posteriormente, se tiene la función llamada “membresías” que a partir de los conjuntos  $A_1, A_2, A_3, A_4, A_5$  estima los valores de membresía correspondiente a cada conjunto, aplicando la función de membresía en campana.

La siguiente función se denomina “Calcular  $V(T)_i$ ” para el proceso de fuzificación y defuzificación.

## RESULTADOS

La diferencia entre la población mexicana del año  $t$  y el año anterior nos da la variación entre  $[t - 1, t]$ . La columna “variación” de la Tabla 2 se constituye en el universo de discurso  $U$ . Sea el valor mínimo de variación de la población entre  $[t - 1, t] = -825,589$  y el máximo valor  $18,621,595$  tenemos un intervalo en el cual fluctúan las variaciones de la población estimada con base a los datos del INEGI 2022. Acorde a la metodología  $U = [V_{min} - D_1, V_{max} - D_2]$ , donde  $D_1 = 0$  y  $D_2 = 0$ , lo que significa que los límites del universo de discurso son exactos debido a que estamos tratando con grandes magnitudes, en este sentido el universo de discurso de la población es  $U = [-825,589; 18,621,595]$ .

Tabla 1: Variables lingüísticas y conjuntos difusos del modelo FTS-P

Conjuntos difusos	Variable lingüística
A1: Muy Bajo crecimiento poblacional en 10 años (MBCP)	MBCP, (-825.589;3.063.847), A1
A2: Bajo crecimiento poblacional en 10 años (BCP)	BCP, (3.063.847; 6.953.284), A2
A3: Moderado crecimiento poblacional en 10 años (MCP)	MCP, (6.953.284; 10.842.721), A3
A4: Alto crecimiento poblacional en 10 años (ACP)	ACP, (10.842.721;14.732.158), A4
A5: Muy Alto crecimiento poblacional (MACP)	MACP, (14.732.158;18.621.595), A5

Fuente: elaboración propia.

El universo de discurso se divide en intervalos iguales. En el presente caso se tiene cinco intervalos que corresponden a las variables lingüísticas definidas en la tabla 1:  $u_1 = [-825,589;3,063,847]$ ,

$u_2 = [3,063,847;6,953,284]$ ,  $u_3 = [6,953,284;10,842,721]$ ,  $u_4 = [10,842,721;14,732,158]$ ,  $u_5 = [14,732,158;18,621,595]$ . Los puntos medios de los intervalos para la estimación del menor error promedio son:  $u_m^1 = 1,119,129$ ,  $u_m^2 = 5,008,566$ ,  $u_m^3 = 8,898,003$ ,  $u_m^4 = 12,787,439$ ,  $u_m^5 = 16,676,876$ . La variable lingüística: “crecimiento decenal de la población mexicana” adopta diferentes valores lingüísticos como se puede observar en la Tabla 1.

Tabla 2: Los datos de la población y los valores de membrecía entre 1900-2020

Año	Población	Variaciones	A1	A2	A3	A4	A5
1895	12491573	0	0	0	0	0	0
1900	13607259	1115686	0.99998814	0.06190219	0.01624314	0.00728705	0.00411268
1910	15160369	1553110	0.8415106	0.07727869	0.01819921	0.00786101	0.00435297
1921	14334780	-825589	0.2091206	0.02854096	0.01046592	0.00536727	0.00325376
1930	16552722	2217942	0.45302389	0.11379693	0.02191867	0.00887199	0.00476053
1940	19653552	3100830	0.20295761	0.21554245	0.02889571	0.01054514	0.00539639
1950	25791017	6137465	0.03819171	0.43967398	0.11600161	0.02211305	0.00892226
1960	34923129	9132112	0.01533557	0.05554435	0.94804071	0.06963099	0.01726416
1970	48225238	13302109	0.00669232	0.01433017	0.04902886	0.79058609	0.0807164
1980	66846833	18621595	0.00325376	0.00536727	0.01046592	0.02854096	0.2091206
1990	81249645	14402812	0.00563519	0.01120425	0.0319459	0.27705226	0.16203858
2000	97483412	16233767	0.0043582	0.0078737	0.0182437	0.07765681	0.8358785
2010	112336538	14853126	0.00527363	0.0102129	0.02742469	0.18985926	0.23115689
2020	126014024	13677486	0.00630071	0.01313195	0.04194023	0.55797897	0.10003657

Fuente: elaboración propia con base a datos del INEGI 2022.

De acuerdo a la metodología, la representación en lenguaje matemático de cada uno de los valores de membrecía que corresponden a cada variación (columna 3 de la Tabla 2) y se despliegan de la columna A1 a A5 es el siguiente:

$$A^{\text{año}} = (f(x_{i,j})/u_1), (f(x_{i,j})/u_2), (f(x_{i,j})/u_3), (f(x_{i,j})/u_4), (f(x_{i,j})/u_5)$$

Donde: año adquiere los valores de 1895 a 2020,  $f(x_{i,j})$  es el valor de membrecía que corresponde a la variación  $i$  y su pertenencia al conjunto  $j$ . Estos conjuntos son calculados con base a la variación de la población con respecto al año anterior y la función de membrecía definida en la ecuación (2) de la metodología.

A partir de la matriz de valores fuzificados se obtienen los valores  $V(T)_i$ , los cuales se suman al valor observado  $X_{i-1}$  para obtener el valor proyectado  $\hat{X}_i$  en la etapa retrospectiva, proceso el cual permite evaluar el ajuste de los datos proyectados *versus* los datos observados (véase Tabla 3).

Tabla 3: Resultados de proyección de la etapa retrospectiva

2020						
Año	Observado		Proyectado		Error	Promedio
	Población	Variación	Población	Variación		
1950	25791017	6137465	23513912	NA	NA	0.03409
1960	34923129	9132112	31725742	8211830	0.02635	
1970	48225238	13302109	43884245	12158503	0.02371	
1980	66846833	18621595	60942232	17057987	0.02339	
1990	81249645	14402812	82333169	21390937	0.08601	
2000	97483412	16233767	94760788	12427619	0.03904	
2010	112336538	14853126	113520854	18760066	0.03478	
2020	126014024	13677486	126528129	13007275	0.00532	
2018						
año	Observado		Proyectado		Error	Promedio
	Población	Variación	Población	Variación		
1950	25791017	6137465	24988466	NA	NA	0.04074
1960	34923129	9132112	32992219	8003753	0.03231	
1970	48225238	13302109	44848417	11856198	0.02998	
1980	66846833	18621595	61531883	16683466	0.02899	
1990	81249645	14402812	82648390	21116507	0.08263	
2000	97483412	16233767	95223174	12574784	0.03753	
2010	112336538	14853126	113778984	18555810	0.03296	

Fuente: elaboración propia.

Como se puede observar en la Tabla 3 y en la Figura 6, la curva en la etapa retrospectiva se ajusta a los datos observados entre 1940 y 2020 con un error promedio de 0,03409. La proyección de datos inicia en 2030, año a partir del cual se aplica la metodología considerando la matriz de valores fuzificados de las variaciones  $R^w(t)$ , o modelo de primer orden para dar los resultados de la etapa prospectiva (2030 a 2100).

También se puede observar en la Tabla 3 los resultados de Argote (2018) que realizó el ejercicio de proyección que no incluía el dato oficial de 2020. Comparando ambos procesos de proyección se puede observar que el error promedio de la proyección disminuye de 0,04074 en 2018 a 0,03409 en 2022, lo cual significa que el dato de 2020 brinda menor incertidumbre a la serie proyectada.

La Tabla 4 muestra la variabilidad de diferentes estimaciones de proyecciones de población total mexicana, realizadas entre 1998 y 2022, conformando así un escenario para el análisis de la precisión del estimador.

Se esperaría que las diferentes estimaciones de la población total para determinado año variaran alrededor de una línea recta, como ocurre con las estimaciones para 2010 y 2020 indicando poca variabilidad entre los diferentes métodos, sin embargo, se observa que a medida que el tiempo de proyección aumenta las estimaciones se alejan y adquieren una curva inestable. Es evidente que el conteo de 2005 define un mínimo en las estimaciones y por el contrario el Censo Nacional de Población y Vivienda de México de 2010 (CNPV) un máximo. El hecho que las curvas cada año se alejen del punto de referencia que puede ser 2010 o 2020 puede indicar el efecto de la incertidumbre que aumenta cuanto más se aleja del presente (véase Figura 4 y Figura 5).

Es interesante observar que al comparar los resultados de la proyección de Argote publicada en 2018 y la de 2022, el periodo proyectado de 2020 a 2050 de 2018 presenta datos por debajo de la curva de 2022, aunque la proyección de 2020 que se realizó en 2018 (126,968,922) es muy cercana al dato oficial (126,014,024) disponible en la actualidad.

También se observa que, a diferencia de la curva de 2018, la proyección de 2022 después de 2030 muestra una curva estable que sube un poco entre 2030-2050 y que posteriormente baja levemente entre 2070 y 2090 para finalmente subir suavemente hacia 2100.

A continuación, se presentará un análisis comparativo de las diferentes proyecciones de la población total mexicana por año en que se realizó el ejercicio de proyección (véase Figura 5). La proyección de la población mexicana a 2030 en 1998 se estima cerca de los 130 millones de habitantes, de acuerdo al conteo de 2005 esta estimación se ajusta y baja considerablemente a cerca de 120 millones para subir posteriormente en las proyecciones de 2009 (126 millones) y 2010 con los datos del CNPV a 136 millones.



Tabla 4: Resultados de proyección de la población total mexicana: comparativo

Etapa	Año	Observado censos y conteos INEGI 2010	Proyección 2000-2050 versión 1998 (1)	Proyección con base al CNPV 2000 (2)	Proyección con base al CONTEO 2005 (3)	Proyección CEPAL 2009 (4)	Proyección con base al CNPV 2010 (5)	Proyección con Series de Tiempo difusas 2016	Proyección con Series de Tiempo difusas 2022
Retrospectiva	1895	12,700,294	NA	NA	NA	NA	NA	NA	NA
	1900	13,607,259	NA	NA	NA	NA	NA	NA	NA
	1910	15,160,369	NA	NA	NA	NA	NA	NA	NA
	1921	14,334,780	NA	NA	NA	NA	NA	NA	NA
	1930	16,552,722	NA	NA	NA	NA	NA	NA	NA
	1940	19,653,552	NA	NA	NA	NA	NA	NA	NA
	1950	25,791,017	NA	NA	NA	27,741,000	NA	23,100,276	23,513,912
	1960	34,923,129	NA	NA	NA	37,877,000	NA	32,297,019	31,725,742
	1970	48,225,238	NA	NA	NA	52,029,000	NA	44,466,262	43,884,245
	1980	66,846,833	NA	NA	NA	69,321,000	NA	61,376,766	60,942,232
Prospectiva	1990	81,249,645	NA	NA	NA	83,906,000	86,313,924	82,608,385	82,333,169
	2000	97,483,412	99,600,000	100,569,263	NA	98,957,000	100,246,535	94,830,105	94,760,788
	2010	112,336,538	112,200,000	111,613,906	108,396,211	110,675,000	113,462,004	113,778,984	113,520,854
	2020	126,014,024	122,100,000	120,639,160	115,762,289	120,099,000	126,441,718	126,968,922	126,528,129
	2030		128,900,000	127,205,386	120,928,075	127,516,000	136,955,288	130,260,380	138,871,373
	2040		132,200,000	130,200,000	122,936,136	132,145,000	145,020,862	133,696,704	142,162,831
	2050		131,600,000	129,592,522	121,855,703	133,341,000	150,622,767	137,632,144	145,599,155
	2060								149,534,595
	2070								153,602,163
	2080								156,033,924
	2090								159,894,284
2100								165,829,009	

1) Proyecciones de población de México 2000-2050 con base a los datos disponible en 1998.

2) Proyecciones de población de México 2000-2050 con base a los datos del Censo de 2000.

3) Proyecciones de Población de México con base al Censo 2005.

4) Proyecciones de Población, Latinoamérica y el Caribe, 2009. Observatorio Demográfico, Año 4, nro 7.

5) Proyecciones de población de México 2010-2050 con base a los datos del Censo de 2010.

6) Proyecciones de población con lógica difusa versión 2018.

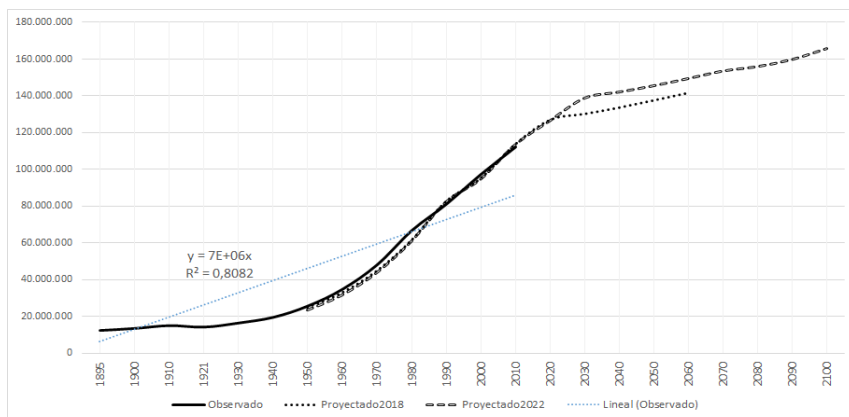
7) Proyecciones de población con Series de Tiempo difusas versión 2022.

NA: No Aplica.

Nota: Se resalta el año 2020 que se incluye como dato en la proyección de 2022.

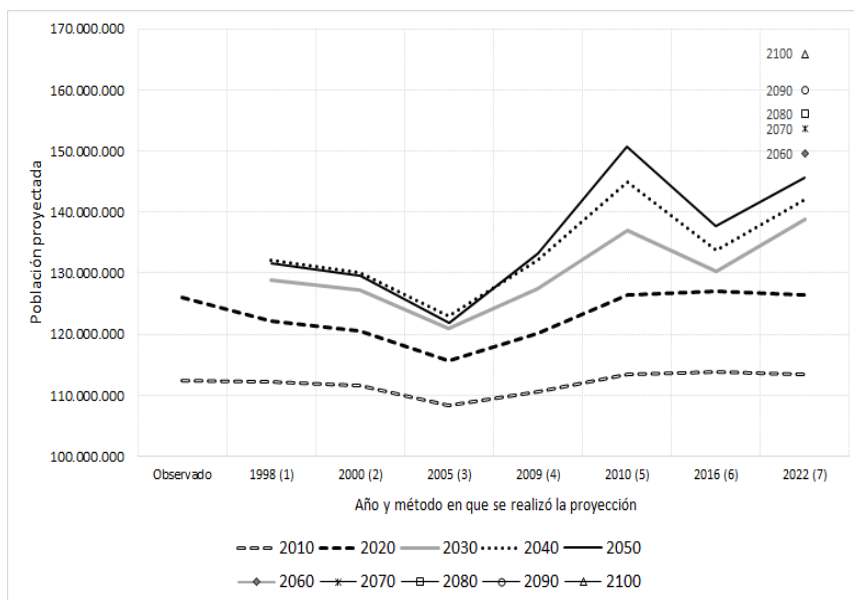
Fuente: elaboración propia.

Figura 4: Población mexicana observada y proyectada 1895 y 2100



Fuente: elaboración propia con base a los datos proyectados con FTS

Figura 5: Diferentes Versiones de la población mexicana proyectada a 2050



(1) Proyecciones de población de México 2000-2050 con base a los datos disponible en 1998; (2) Proyecciones de población de México 2000-2050 con base a los datos del Censo de 2000; (3) Proyecciones de población de México con base al Censo 2005; (4) Proyecciones de Población, Latino América y el Caribe, 2009. Observatorio Demográfico, Año 4, nro. 7; (5) Proyecciones de población de México 2010-2050 con base a los datos del Censo de 2010; (6) Proyecciones de población de México 2010-2050 estimada con series de tiempo difusas versión 2016 (7) Proyecciones de población de México 2020-2100 estimada con series de tiempo difusas versión 2022.

Finalmente, en 2018 la autora realiza una estimación también cercana a 130 millones, mientras que en la presente investigación 2022 se brinda un resultado de 138 millones muy cercano a la estimación con base del censo de 2010. Es decir, podríamos establecer un intervalo de confianza para 2030 entre [120.928.075 a 138.871.373].

Para la serie de estimaciones de 2040 se mantienen las tendencias de 2030. Un mínimo para 2005 y un máximo para 2010. La estimación vuelve a bajar un poco con STD en 2018 y en 2022 que considera el dato de 2020, vuelve a subir, estableciéndose un intervalo entre [122.936.136 a 145.020.862].

Para 2050 se evidencia el mismo comportamiento de 2030 y 2040 con una estimación de la población total mexicana entre [121.855.703 a 150.622.767].

## CONCLUSIONES

Tomando como referencia la investigación de la autora publicada en 2018 que proyecta la población mexicana con Series de Tiempo Difusas entre 2020-2050 con resultados coherentes de acuerdo al análisis comparativo con otras estimaciones oficiales del país, el presente artículo es la continuación de dicha investigación considerando los siguientes frentes: primero, en 2022 se cuenta con el dato oficial de 2020 el cual se puede evaluar con los resultados de 2018 y usar como dato para proyectar nuevamente y evaluar el impacto, segundo, con la experiencia pasada se crea un algoritmo a la medida que encapsula el método de STD acorde a las necesidades, de tal manera que se pueden realizar los procesos de forma automática lo cual reduce el tiempo de cómputo y tercero, ya que se realizan dos ejercicios de proyección con STD en dos momentos en el tiempo (2018 y 2022) se cuenta con mayor evidencia para validar los resultados de la proyección con STD que incorpora el comportamiento no lineal a través de una lógica borrosa.

Los resultados actualizados de la proyección de la población mexicana que incluye el dato de 2020 muestran resultados coherentes y brindan un escenario factible de la dinámica poblacional a 2100. En términos del error, la proyección realizada en 2022 tiene un menor error promedio (0,03409) que la proyección realizada en 2018 (0,04074), lo cual significa que la inclusión del dato 2020 redujo la incertidumbre de la serie de datos. Adicionalmente, el método actualizado y las características de la investigación permitieron proyectar a un mayor horizonte.

La comparación de los resultados de la proyección con STD con otras estimaciones oficiales en diferentes momentos en el tiempo muestran la robustez del método. Para 2010 y 2020 la estimación con STD se acerca al promedio de todos los puntos de comparación, lo cual indicaría que el método estaría reflejando la media del comportamiento del fenómeno en estudio. Para 2030, 2040 y 2050 si bien los resultados de la autora en 2018 también tendían al promedio de las estimaciones, la inclusión del dato de 2020 en el ejercicio de 2022 muestra un incremento de las estimaciones capturando así el efecto del CNPV de 2010. Para el futuro de 2060 a 2100 no se encontraron referentes comparativos, sin embargo, las estimaciones muestran una tendencia coherente.

La presente investigación también permitió validar el algoritmo para proyectar con STD considerando una función de membresía en campana. Si bien, es factible encontrar código desarrollado en github (plataforma compartida para proyectos de generación de código), la desventaja de estas librerías o funciones es que son una caja negra para quien está modelando determinado fenómeno, por lo que el desarrollo a medida es la mejor opción para satisfacer las necesidades específicas. En este sentido, el algoritmo creado en R es una herramienta que permite la simulación de diversos escenarios e incrementar la capacidad de cómputo.

Finalmente, como se pudo evidenciar en el marco teórico, existe una abundante investigación en el área de proyección de datos con STD, los avances en el uso de la técnica se han realizado en la búsqueda de la precisión, optimizando procesos de la metodología. En este contexto, la presente investigación aporta en el análisis en profundidad de la metodología para el caso de la población total mexicana y en un análisis de sensibilidad de las estimaciones que no se había realizado con anterioridad.

## REFERENCIAS BIBLIOGRÁFICAS

- Abbasov, A., & Mamedova, M., 2003, "Application of fuzzy time series to population forecasting", in *Vienna University of Technology*, 1, 545–552.
- Aladag, S., Aladag, C. H., Mentis, T. & Egrioglu, E., 2012, "A new seasonal fuzzy time series method based on the multiplicative neuron model and SARIMA", in *Journal of Mathematics and Statistics*, 41(3), 145–163.
- Alho, J., Alders, M., Crujisen, H., Keilman, N., Nikander, T., & Pham, D. Q., 2006, "New forecast: Population decline postponed in Europe", in *Statistical Journal of the United Nations Economic Commission for Europe*, 23(1), 1-10.

Aqil Burney, S.M. y Akhter Raza, S., 2005, “Time Series Forecasting with SA-RIMA Model: A Case Study Using the Natural Gas Demand Series”, in *Karachi University Journal of Science* Vol. 33 (1 & 2) July — December 2005, 31-35.

Argote Cusi, M.L. y Parra Bernal, L.D., 2020, *Global Entrepreneurship Analytics: Using GEM Data*. Routledge.

Argote Cusi, M.L., 2018, “El uso de lógica difusa en proyecciones de población: el caso de México”, en *Papeles de población*, 24(95), 273-301.

Argote Cusi, M.L., 2016, “Uso de la lógica difusa en proyecciones de población”. Ponencia enviada a la XIII Reunión Nacional de Investigación Demográfica en México realizada entre el 22 y 24 de junio, Universidad Nacional Autónoma de México, ciudad de México, 2016.

Arun Kumar, K. E., Kalaga, D. V., Kumar, C. M. S., Kawaji, M., & Brenza, T. M., 2022, “Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting Covid-19 trends”, in *Alexandria Engineering Journal*, 61(10), 7585-7603.

Bas, E., Uslu, V. R., Aladag, C., Yolcu, U., & Egrioglu, E., 2014, “A modified genetic algorithm for forecasting fuzzy time series”, in *Applied Intelligence*, 41, 453–463.

Burney, S. A., Ali, S. M. and Khan, M. S., 2018, “A novel high order Fuzzy Time Series forecasting method with higher accuracy rate”, in *IJCSNS*, 18(5), 9.

CEPAL, 2009, “Proyección de Población”, en *Observatorio Demográfico: América Latina y el Caribe*, núm. 7, abril/2009.

Chen, S.M., 1996, “Forecasting Enrollments Based on Fuzzy Time Series”, in *Fuzzy Sets and Systems*. 81, 311–319.

Chen, S. M., & Tanuwijaya, K., 2011, “Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques”, in *Expert Systems with Applications*, 38, 15425–15437.

Chen, S. M., & Hsu, C. C., 2004, “A new method to forecast enrollments using fuzzy time series”, in *International Journal of Applied Science and Engineering*, 2(3), 234-244.

Chen, S. M. and Hwang, J. R., 2000, “Temperature prediction using fuzzy time series”, in *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, 30: 263-275.

Egrioglu, S., Bas, E., Aladag, C. H., & Yolcu, U., 2016, “Probabilistic fuzzy time series method based on artificial neural network”, in *American Journal of Intelligent Systems*, 62(2), 42–47.

Ghosh, H., Chowdhury, S., & Prajneshu, S., 2015, “An improved fuzzy time series method of forecasting based on L-R fuzzy”, in *Journal of Applied Statistics*, 43(6), 1128–1139.

- Gorbatiuk, K., Hryhoruk, P., Proskurovych, O., Rizun, N., Gargasas, A., Raupe-  
lienè, A., & Munjishvili, T., 2021, "Application of fuzzy time series forecasting  
approach for predicting an enterprise net income level", in E3S Web of Conferen-  
ces (Vol. 280). EDP Sciences.
- Huang, K., 2001, "Effective lengths of intervals to improve forecasting in fuzzy  
time series", in Fuzzy sets and systems, 123(3), 387-394.
- INEGI, 2022, *Demografía y Sociedad*, Disponible en [https://www.inegi.org.mx/  
temas/estructura/](https://www.inegi.org.mx/temas/estructura/)
- Jang, J. S. R., Sun, C. T., Mizutani, E., 1997, "Neuro-fuzzy and soft computing-a  
computational approach to learning and machine intelligence", in *IEEE Transac-  
tions on automatic control*, 42(10), 1482-1484.
- Khan, M. S., Burney, S. A., & Ali, S. M., 2018, "A Novel High Order Fuzzy Time  
Series Forecasting Method with Higher Accuracy Rate", in *IJCSNS International  
Journal of Computer Science and Network Security*, vol. 18 No. 5.
- Kopcsó, D., & Pachamanova, D., 2018, "Case Article—Managing Staffing In-  
efficiencies Using Analytics (B): Business Value in Predictive and Prescriptive  
Analytics Models", in *INFORMS Transactions on Education*, 19(1), 43-47.
- Li, S. T., & Chen, Y. P., 2004, "Natural partitioning-based forecasting model for  
fuzzy time-series", in 2004 IEEE International Conference on Fuzzy Systems  
(IEEE Cat. No. 04CH37542) (Vol. 3, pp. 1355-1359). IEEE.
- Li, Y., & Pan, Y., 2022, "A novel ensemble deep learning model for stock predic-  
tion based on stock prices and news", in *International Journal of Data Science and  
Analytics*, 13(2), 139-149.
- Mustapha, F. Z., Haruna, A. A., & Muhammad, U. T., 2020, "An Overview of  
Artificial Intelligence", in *Journal of Applied Sciences & Environmental Sustain-  
ability*, 6(12), 60-74.
- Nosratabadi, S., Mosavi, A., Duan, P., Ghamisi, P., Filip, F., Band, S. S., & Gando-  
mi, A. H., 2020, "Data science in economics: comprehensive review of advanced  
machine learning and deep learning methods", in *Mathematics*, 8(10), 1799.
- Ordorica Mellado, M., 2004, "Pronóstico de las defunciones por medio de los mo-  
delos autorregresivos integrados de promedios móviles", en *Papeles de población*,  
10(42), 249-264.
- Panigrahi, S., & Behera, H. S., 2020, "A study on leading machine learning tech-  
niques for high order fuzzy time series forecasting", in *Engineering Applications  
of Artificial Intelligence*, 87, 103245.
- Pan, D., Yang, J., Zhou, G. y Kong, F., 2020, The influence of COVID-19 on  
agricultural economy and emergency mitigation measures in China: A text min-  
ing analysis. *PLoS ONE* 15(10): e0241167. [https://doi.org/10.1371/journal.  
pone.0241167](https://doi.org/10.1371/journal.pone.0241167)

- Raftery A.E. y Ševčíková H., 2021, “Probabilistic population forecasting: Short to very long-term”, in *International Journal of Forecasting*, <https://doi.org/10.1016/j.ijforecast.2021.09.001>.
- Rana, A.K., 2020, “Fish Production Forecasting in India Using Nested Interval Based Fuzzy Time Series Model”, in *International Journal of Recent Technology and Engineering*. 8(6), 5534-5537.
- Sasu, A., 2010, “An application of fuzzy time series to the Romanian population”, in *Bulletin of the Transilvania University of Brasov Vol. 3*, 52.
- Souma, W., Vodenska, I., & Aoyama, H., 2019, “Enhanced news sentiment analysis using deep learning methods”, in *Journal of Computational Social Science*, 2(1), 33-46.
- Song, Q., & Chissom, B. S., 1993, “Fuzzy time series and its models”, in *Fuzzy Sets and Systems*, 54(3), 269–277.
- Song Q. and Chissom, B.S., 1994, “Forecasting enrollments with fuzzy time series - part II”, *Fuzzy Sets and Systems* 62, 1-8.
- Sullivan, J., & Woodall, W. H., 1994, A comparison of fuzzy forecasting and Markov modeling. *Fuzzy Sets and Systems*, 64(3), 279-293.
- Vovan, T., 2019, “An Improved Fuzzy Time Series Forecasting Model Using Variations of Data”, in *Fuzzy Optimization and Decision Making*. 18, 151-173.
- Wang, S., Li, C., & Lim, A., 2019, Why are the ARIMA and SARIMA not sufficient. arXiv preprint arXiv:1904.07632.
- Yusuf, S. M., Mohammad, A., & Hamisu, A. A., 2017, “A novel two-factor high order fuzzy time series with applications to temperature and futures exchange forecasting”, in *Nigerian Journal of Technology*, 36(4), 1124-1134.
- Zadeh, L. A., 1973, “Outline of a new approach to the analysis of complex systems and decision processes”, in *Systems, Man and Cybernetics*, IEEE Transactions on, (1), 28-44.

## RESUMEN CURRICULAR DE LA AUTORA Y EL AUTOR

*Milenska Linneth Argote Cusi*

Co-fundadora de BI&DE SAS y de la Fundación ShareYourKnowHow. Conferencista. Miembro de Data Science Specialization of Johns Hopkins University. Maestra en Estudios de Población, FLACSO, México. Ingeniera de Sistemas, Escuela Militar de Ingeniería, Bolivia. Diplomada en Ciberseguridad, Universidad Piloto de Colombia. Docente Seminario de Investigación, Universidad Santo Tomas, Colombia. Miembro del Entrepreneurship Group, MinCiencias, Colombia. Tiene más de quince años de experiencia en investigación, consultoría y docencia en diferentes instituciones privadas y públicas de América Latina. Consultorías recientes: De-

sarrollo de un Sistema de Gestión de Seguridad de la Información para una empresa del sector metalmecánico (2023), Inteligencia Artificial para el pronóstico de datos (2022), Programa de Certificación en Data Analytics a Puerto Rico Manufacturing Extension (PRIMEX) 2020, Programa de Fortalecimiento de Capacidades en Data Analytics en empresas bogotanas, MINTIC-ViveLAB (2019). Publicaciones más recientes: Proyección de la población mexicana usando Fuzzy Times Series (AI): horizonte 2100 (2023) Colombian agricultural sector's early estimator of Gross Domestic Production post-pandemic Covid-19 using Google News and Google Trends (2022), Global Entrepreneurship Analytics: using GEM Data, Routledge (2020), Análisis de brechas tecnológicas en el sector metalmecánico (2019). Líneas de investigación: Ciencia de Datos, Inteligencia Artificial, Talento Humano, Ciberseguridad.

Dirección electrónica: margotecusi@gmail.com

Registro ORCID: <http://orcid.org/0000-0003-4448-1347>

*León Darío Parra Bernal*

Economista, Magíster en Estudios de Población, FLACSO, sede México, Estudios de Doctorado en Economía, UNAM, México, se ha desempeñado como docente durante veinte años en el área empresarial y Data Analytics en universidades públicas y privadas de Latinoamérica. Ha liderado varios proyectos de investigación y consultoría con diferentes organismos internacionales como ONUSIDA, PNUD, UNFPA y el Global Entrepreneurship Monitor (GEM) en las áreas de desarrollo regional, brechas tecnológicas, Big Data y Data Analytics a nivel global. Actualmente se desempeña como profesor asociado en la Universidad EAN, asesor técnico del Sistema de Estadística Nacional (SEN) en modernización tecnológica y empresario en el área de análisis de datos y transformación digital.

Dirección electrónica: [ldparra@universidadean.edu.co](mailto:ldparra@universidadean.edu.co)

Registro ORCID: <http://orcid.org/0000-0002-1509-1262>