

*Trabajo Fin de Máster*

“Predicción de la mortalidad mediante Redes  
Neuronales Recurrentes de tipo Long Short-Term  
Memory”

---

**Beatriz Bolado Contreras**

Tutor/es

José Miguel Rodríguez-Pardo del Castillo

Jesús Ramón Simón del Potro

15 de Julio de 2020



Esta obra se encuentra sujeta a la licencia Creative Commons  
**Reconocimiento – No Comercial – Sin Obra Derivada**

Esta tesis es propiedad del autor. No está permitida la reproducción total o parcial de este documento sin mencionar su fuente. El contenido de este documento es de exclusiva responsabilidad del autor, quien declara que no se ha incurrido en plagio y que la totalidad de referencias a otros autores han sido expresadas en el texto.

En caso de obtener una calificación igual o superior a 9.0 (Sobresaliente), autorizo la publicación de este trabajo en el centro de Documentación de la Fundación Mapfre.

Sí, autorizo a su publicación.

No, desestimo su publicación.

Firmado:

A handwritten signature in black ink, enclosed within a faint, hand-drawn oval shape.

*A mi familia y a los que se han convertido en ella,  
en especial a mis padres por su incondicional apoyo.*

## **Resumen**

Los modelos tradicionales de predicción de la mortalidad han gozado de una amplia aceptación entre los expertos durante los últimos veinte años. No obstante, entre todas sus bondades también se destacan ciertas deficiencias como la invariabilidad en el tiempo de los parámetros relacionados con la edad. El objetivo de este trabajo es el logro de una mejora de la predicción de la mortalidad implementando modelos que permitan una gran capacidad computacional, como son las redes neuronales artificiales. La introducción de las redes neuronales recurrentes de memoria a corto-largo plazo (LSTM) permite crear secuencias de datos y relacionarlas, recordando información durante largos períodos de tiempo y reconociendo patrones de comportamiento de forma hábil. Este tipo de redes de memoria permiten, mediante su innovadora estructura, superar las limitaciones de los modelos tradicionales al ser capaces de clasificar, generar y predecir secuencias, ofreciendo por ello una predicción de la mortalidad más precisa.

Palabras Clave: Redes neuronales artificiales; Mortalidad; Memoria a corto-largo plazo.

## **Abstract**

Traditional mortality prediction models have been widely accepted among experts over the past twenty years. However, among their many advantages, there are also certain deficiencies, such as the invariance of age-related parameters over time. The aim of this work is to improve the prediction of mortality by implementing models that allow a great deal of computational capacity, such as artificial neuronal networks. The introduction of long short-term memory (LSTM) recurrent neuronal networks allows to create data sequences and to relate them, remembering information during long periods of time and recognizing behavior patterns in a skillful way. This type of memory network allows, through its innovative structure, to overcome the limitations of traditional models by being able to classify, generate and predict sequences, thus offering a more accurate prediction of mortality.

Keywords: Artificial neural networks; Mortality; Long short-term memory.

# ÍNDICE DE CONTENIDOS

CAPÍTULO 1. INTRODUCCIÓN.....	8
CAPÍTULO 2. MODELO DE LEE CARTER .....	10
CAPÍTULO 3. REDES NEURONALES.....	15
3.1. Redes neuronales biológicas y redes neuronales artificiales .....	15
3.2. Definiciones.....	17
3.3. Propiedades.....	18
3.4. Morfología de una neurona .....	19
3.5. Tipos de funciones de activación .....	21
3.6. Topología de las redes neuronales.....	24
3.7. Entrenamiento de una red neuronal .....	27
3.8. Algoritmo de retro propagación .....	29
CAPÍTULO 4. REDES NEURONALES RECURRENTES .....	33
CAPÍTULO 5. LONG-SHORT TERM MEMORY .....	37
CAPÍTULO 6. CASO PRÁCTICO: PREDICCIÓN DE LA MORTALIDAD.....	44
6.1. Aplicación del Modelo de Lee-Carter .....	46
6.2. Aplicación de la red neuronal recurrente de tipo LSTM.....	56
6.2.1. Elección de los hiper parámetros. ....	57
6.2.2. Resultados .....	64
6.3. Análisis conjunto del tramo de edades 80-100.....	71
CAPÍTULO 7. CONCLUSIONES.....	77
BIBLIOGRAFÍA.....	79
ANEXO A.....	81
A.1. Análisis en la población femenina para el año 2007 .....	81
A.2. Análisis en la población femenina para el año 2011 .....	83
A.3. Análisis en la población femenina para el año 2016. ....	85
ANEXO B .....	89
B.1. Análisis en la población masculina para el año 2007 .....	89
B.2. Análisis en la población masculina para el año 2011 .....	91
B.3. Análisis en la población masculina para el año 2016 .....	93

## ÍNDICE DE TABLAS

Tabla 1. Ejemplo del formato largo aplicado.....	45
Tabla 2. Evaluación de la predicción mediante LC para el género femenino.....	53
Tabla 3. Evaluación de la predicción mediante LC para el género masculino. ....	56
Tabla 4. Pruebas de hiper parámetros para el género femenino. ....	60
Tabla 5. Pruebas de hiper parámetros para el género masculino. ....	61
Tabla 6. Evaluación de la predicción mediante LSTM para el género femenino. ....	68
Tabla 7. Evaluación de la predicción mediante LSTM para el género masculino.....	71
Tabla 8. Zoom en el tramo 80-99 años para el género femenino.....	74
Tabla 9. Zoom en el tramo 80-99 años para el género masculino. ....	75

## ÍNDICE DE FIGURAS

Figura 1. Morfología de una neurona humana. ....	16
Figura 2. Esquema básico de una neurona. ....	20
Figura 3. Representación de una función lineal. ....	22
Figura 4. Representación de una función logística. ....	23
Figura 5. Representación de una función hiperbólica tangente. ....	23
Figura 6. Representación de una función ReLU. ....	24
Figura 7. Diagrama de una red neuronal de una sola capa. ....	25
Figura 8. Diagrama de una red neuronal con una capa oculta. ....	26
Figura 9. Diagrama de una red neuronal con dos capas ocultas. ....	26
Figura 10. Esquema sobre el entrenamiento de una red neuronal. ....	29
Figura 11. Representación gráfica del descenso del gradiente. ....	31
Figura 12. Comparación de una red neuronal básica con una recurrente. ....	34
Figura 13. Despliegue de una red neuronal recurrente. ....	34
Figura 14. Esquema básico de una red LSTM. ....	38
Figura 15. Representación de la capa de olvido en una red LSTM. ....	39
Figura 16. Representación de una puerta de entrada en una red LSTM. ....	40
Figura 17. Representación de operaciones puntuales en una red LSTM. ....	41
Figura 18. Representación de una capa de salida en una red LSTM. ....	42
Figura 19. Mapa de calor de la log-mortalidad para el género femenino en España. ...	45
Figura 20. Mapa de calor de la log-mortalidad para el género masculino en España. ..	46
Figura 21. Representación del parámetro $\alpha$ para el género femenino. ....	47
Figura 22. Representación del parámetro $\alpha$ para el género masculino. ....	47
Figura 23. Representación del parámetro $\beta$ para el género femenino. ....	48
Figura 24. Representación del parámetro $\beta$ para el género masculino. ....	48
Figura 25. Representación del parámetro $\kappa$ proyectado para el género femenino. ....	49
Figura 26. Representación del parámetro $\kappa$ proyectado para el género masculino. ....	49
Figura 27. Mortalidad observada y predicha en el año 2007 para el género femenino mediante LC. ....	50
Figura 28. Log-Mortalidad observada y predicha en el año 2007 para el género femenino mediante LC. ....	50
Figura 29. Mortalidad observada y predicha en el año 2011 para el género femenino mediante LC. ....	51
Figura 30. Log-Mortalidad observada y predicha en el año 2011 para el género femenino mediante LC. ....	51
Figura 31. Mortalidad observada y predicha en el año 2016 para el género femenino mediante LC. ....	52
Figura 32. Log-Mortalidad observada y predicha en el año 2016 para el género femenino mediante LC. ....	52
Figura 33. Mortalidad observada y predicha en el año 2007 para el género masculino mediante LC. ....	53
Figura 34. Log-Mortalidad observada y predicha en el año 2007 para el género masculino mediante LC. ....	54

Figura 35. Mortalidad observada y predicha en el año 2011 para el género masculino mediante LC.....	54
Figura 36. Log-Mortalidad observada y predicha en el año 2011 para el género masculino mediante LC.....	55
Figura 37. Mortalidad observada y predicha en el año 2016 para el género masculino mediante LC.....	55
Figura 38. Log-Mortalidad observada y predicha en el año 2016 para el género masculino mediante LC.....	56
Figura 39. Representación del sobreajuste y subajuste.....	63
Figura 40. Representación de la regla de parada temprana en el modelo elegido. ....	64
Figura 41. Mortalidad observada y predicha en el año 2007 para el género femenino mediante LSTM. ....	65
Figura 42. Log-Mortalidad observada y predicha en el año 2007 para el género femenino mediante LSTM. ....	65
Figura 43. Mortalidad observada y predicha en el año 2011 para el género femenino mediante LSTM. ....	66
Figura 44. Log-Mortalidad observada y predicha en el año 2011 para el género femenino mediante LSTM. ....	66
Figura 45. Mortalidad observada y predicha en el año 2016 para el género femenino mediante LSTM. ....	67
Figura 46. Log-Mortalidad observada y predicha en el año 2016 para el género femenino mediante LSTM. ....	67
Figura 47. Mortalidad observada y predicha en el año 2007 para el género masculino mediante LSTM. ....	68
Figura 48. Log-Mortalidad observada y predicha en el año 2007 para el género masculino mediante LSTM. ....	69
Figura 49. Mortalidad observada y predicha en el año 2011 para el género masculino mediante LSTM. ....	69
Figura 50. Log-Mortalidad observada y predicha en el año 2011 para el género masculino mediante LSTM. ....	70
Figura 51. Mortalidad observada y predicha en el año 2016 para el género masculino mediante LSTM. ....	70
Figura 52. Log-Mortalidad observada y predicha en el año 2016 para el género masculino mediante LSTM. ....	71
Figura 53. Comparación en el tramo de edades 80-100 para el año 2007. ....	72
Figura 54. Comparación en el tramo de edades 80-100 para el año 2011. ....	73
Figura 55. Comparación en el tramo de edades 80-100 para el año 2016. ....	73



## CAPÍTULO 1. INTRODUCCIÓN

Los datos demográficos empíricos muestran que las tasas de mortalidad han decrecido de forma constante desde la segunda mitad del siglo XX en los países más desarrollados, suponiendo un considerable aumento de la esperanza de vida de forma generalizada, que se acentúa especialmente en las edades más avanzadas. En las Ciencias Actuariales, el estudio de la mortalidad, concretamente su modelización y predicción, es una cuestión de alta importancia. Esto es debido a que el futuro aumento de la longevidad supone un problema tanto para las empresas aseguradoras, obligándolas a tener que reevaluar el precio de sus productos ofertados, así como sus inversiones en instrumentos financieros, como para los sistemas de pensiones públicos, llevando a tener que realizar una reestructuración de los planes de pensiones y el sistema público de seguridad social. Por lo tanto, se podría decir que la dualidad de la mortalidad y la longevidad es una cuestión que afecta a diferentes disciplinas como la economía, la demografía y las ciencias sociales con igual importancia.

En cuanto a la modelización y predicción de la mortalidad, el conjunto de modelos de Lee-Carter, es decir, su primera formulación ideada en 1992 así como todas sus reconocidas variaciones, han adquirido gran aceptación por los expertos a lo largo de los últimos 20 años. Su sencillez y parsimonia han permitido describir la mortalidad a lo largo del tiempo y predecir su comportamiento a futuro con cierta estabilidad en todas las edades, aunque numerosos inconvenientes y limitaciones son reconocidos por diversos autores, destacando principalmente su invariabilidad en el tiempo.

Sin embargo, nuevas técnicas computacionales han ido adquiriendo gran popularidad en los últimos años, como son el Deep Learning y Machine Learning. Estas, utilizan nuevos algoritmos que destacan en detectar patrones que no son fácilmente identificables o reconocidos mediante el uso de métodos más tradicionales, como pueden ser regresiones o modelos estocásticos. En el campo de la demografía, su utilización es aún temprana, pero está adquiriendo popularidad entre los investigadores debido a que las predicciones se basan más en el propio dato y no tanto en hipótesis.

En este trabajo se propone utilizar una técnica de Deep Learning como son las redes neuronales, frente a uno de los modelos estocásticos más conocidos como es el de Lee Carter tradicional de 1992. El objetivo es intentar aumentar la capacidad predictiva y la

precisión en la proyección de la mortalidad a futuro, evaluando si este método puede ser una alternativa para predecir la mortalidad. En concreto, se utiliza una variación de las redes neuronales tradicionales, como son las redes neuronales recurrentes de tipo Long Short-Term Memory o de Memoria a Corto-Largo plazo, creadas en 1997 por Hochreiter y Schmidhubter. Este tipo de red neuronal está estructurada de forma que es capaz de crear secuencias de datos y relacionarlas, permitiendo recordar la información durante largos periodos de tiempo y reconocer patrones de comportamiento tanto a largo como a corto plazo.

En términos de demografía, las redes neuronales LSTM permiten proyectar la mortalidad teniendo en cuenta toda la influencia de la mortalidad pasada. Por ello, evita que los comportamientos pasados se desvanezcan, utilizando toda la información que considere relevante a lo largo del tiempo, permitiendo una visión de la tendencia de una serie temporal más global y precisa.

El trabajo se organiza de la siguiente forma. El Capítulo 2 se centra en el modelo de Lee Carter y sus variaciones más importantes, explicando su metodología esencial, así como sus bondades y deficiencias. En el Capítulo 3 se trata de explicar los conceptos básicos de las redes neuronales, empezando por su origen y propiedades indispensables, para pasar a temas más específicos como las topologías de redes, los diferentes tipos de aprendizajes y funciones de activación, y el conocido algoritmo de retro propagación. El Capítulo 4 focaliza en las redes neuronales recurrentes y una explicación más concreta de este tipo topología, para dar paso al Capítulo 5, donde se tratarán todas las cuestiones claves de la red neuronal utilizada en el trabajo, Long Short-Term Memory. Finalmente, en el Capítulo 6 se realiza de forma práctica una comparación de la predicción de la mortalidad a 1, 5 y 10 años para la población española mediante una metodología tradicional, como es el modelo de Lee Carter, y una red neuronal recurrente LSTM. Las conclusiones obtenidas a lo largo del trabajo quedan reflejadas en el Capítulo 7, y al final de éste se pueden encontrar los anexos más relevantes para una mayor ampliación de la información.

## CAPÍTULO 2. MODELO DE LEE CARTER

El modelo de Lee Carter fue desarrollado por primera vez en 1992 por Ronald D. Lee y Lawrence R. Carter. El objetivo principal era modelizar y predecir la mortalidad de Estados Unidos, así como evaluar a posteriori su precisión histórica (Lee, Ronald D. & Carter, 1992). Para ello, los autores utilizan datos anuales específicos entre los años 1933 y 1987 para cada edad de la población estadounidense. En pocas palabras, Lee y Carter intentan resumir el plano periodo-edad de las tasas de mortalidad logarítmicas en términos de vectores  $\alpha_x$  y  $\beta_x$  a lo largo de la dimensión de la edad y  $\kappa_t$  a lo largo de la dimensión del tiempo.

El modelo tiene la siguiente expresión principal:

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

que también puede ser escrito como:

$$m_{x,t} = \exp(\alpha_x + \beta_x \kappa_t + \epsilon_{x,t}) = e^{\alpha_x} e^{\beta_x \kappa_t} e^{\epsilon_{x,t}}$$

donde  $m_{x,t}$  es la tasa central de mortalidad a la edad  $x$  en el año del calendario  $t$ ,  $\alpha_x$  y  $\beta_x$  son parámetros o constantes dependientes de la edad y  $\kappa_t$  un índice que depende del tiempo y representa la tendencia temporal subyacente. Específicamente,  $\alpha_x$  es la media de las tasas de mortalidad logarítmicas a lo largo del tiempo, de tal forma que la exponencial del parámetro representa la forma general del perfil de mortalidad por edades, y  $\beta_x$  es la sensibilidad de la tasa de mortalidad logarítmica a la tendencia temporal representada por  $\kappa_t$  para la edad  $x$ . Finalmente,  $\epsilon_{x,t}$  es el término de error y representa los efectos no capturados por el modelo. Se caracteriza por ser una variable aleatoria independientemente e idénticamente distribuida (iid) que sigue una distribución  $N(0, \sigma^2)$ .

Para asegurar que las probabilidades de fallecimiento  $q_{x,t}$  son  $\leq 1$ , el modelo de Lee Carter modela las tasas de mortalidad logit de la siguiente forma:

$$\ln\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}$$

que deriva en:

$$q_{x,t} = \frac{\exp(\alpha_x + \beta_x \kappa_t)}{1 + \exp(\alpha_x + \beta_x \kappa_t)} + \epsilon_{x,t}$$

La estructura es invariante ante las siguientes transformaciones en los parámetros para cualquier constante  $c$ :

$$\{\alpha_x, \beta_x, \kappa_t\} \rightarrow \{\alpha_x, \beta_x/c, c\kappa_t\}$$

$$\{\alpha_x, \beta_x, \kappa_t\} \rightarrow \{\alpha_x - c\beta_x, \beta_x, \kappa_t + c\}$$

Por ello, para evitar problemas de identificabilidad en los parámetros, el modelo requiere las siguientes restricciones:

$$\sum_x \beta_x = 1$$

$$\sum_t \kappa_t = 0$$

Para la estimación de los parámetros se realiza un proceso en dos pasos. En primer lugar, se estima  $\alpha_x$  promediando las tasas logarítmicas a lo largo del tiempo y posteriormente,  $\beta_x$  y  $\kappa_t$  mediante descomposición singular de valores (SVD) de los residuos. Por tanto, el parámetro  $\alpha_x$  es estimado de forma iterativa, empezando en:

$$\hat{\alpha}_x^1 = \frac{\sum \ln(m_{x,t})}{T}$$

donde  $T$  es el número de años  $t$ .

Después, los parámetros  $\beta_x$  y  $\kappa_t$  se estiman como los primeros vectores singulares derecho e izquierdo en la descomposición singular de valores de la matriz de valores:

$$[\ln(m_{x,t}) - \hat{\alpha}_x^1]$$

Siendo  $\hat{\kappa}_t^{SVD}$  y  $\hat{\beta}_x^{SVD}$  el primer vector singular derecho e izquierdo respectivamente.

Por lo tanto, la solución cuando  $\sum_x \hat{\beta}_x = 1$  es:

$$\hat{\beta}_x = \frac{\hat{\beta}_x^{SVD}}{\sum_x \hat{\beta}_x^{SVD}}$$

y ello deriva en que el  $\hat{\kappa}_t^l$  inicial  $\hat{\kappa}_t^l = \hat{\kappa}_t^{SVD} \sum_x \hat{\beta}_x^{SVD}$  es tal que:

$$\hat{\beta}_x \hat{\kappa}_t^l = \hat{\kappa}_t^{SVD} \hat{\beta}_x^{SVD}$$

Finalmente,  $\kappa_t$  son reajustados para que:

$$\sum_x D_{x,t} = \sum_x \hat{D}_{x,t} \quad \forall t$$

donde  $\hat{D}_{x,t} = N_{x,t} \exp(\alpha_x + \beta_x \kappa_t + \epsilon_{x,t})$ , siendo  $\hat{D}_{x,t}$  el número de fallecimientos en la edad  $x$  al tiempo  $t$ , y  $N_{x,t}$  la exposición al riesgo de fallecimiento. Esto nos da el nuevo ajuste  $\hat{\kappa}_t^2$ , y finalmente:

$$\hat{\kappa}_t = \hat{\kappa}_t^2 - \overline{\hat{\kappa}_t^2}$$

$$\hat{\alpha}_x = \hat{\alpha}_x^l + \hat{\beta}_x \overline{\hat{\kappa}_t^2}$$

Una vez que los valores de  $\kappa_t$  han sido estimados, el último paso en el método de Lee Carter es encontrar un modelo ARIMA mediante el uso de la metodología Box-Jenkins. Esto permite el pronóstico de  $\kappa_t$  pasado el último año observado.

Asumiendo que  $\alpha_x$  y  $\beta_x$  permanecen constantes en el tiempo, los autores probaron varias especificaciones para concluir que un paseo aleatorio con deriva también conocido como ARIMA (0,1,0) es el más apropiado para predecir la mortalidad:

$$\kappa_t = \kappa_{t-1} + \theta + \epsilon_t$$

donde  $\theta$  es el parámetro de deriva y los errores  $\epsilon_t$  son iid  $N(0, \sigma^2)$ .

El modelo de Lee Carter tiene reconocidas numerosas bondades, y entre ellas se destaca la combinación de un modelo demográfico parsimonioso con la utilización de una técnica estadística como son las series temporales. Otra ventaja reseñable es que las predicciones están basadas en tendencias a largo plazo donde se proporcionan intervalos de confianza probabilísticos, permitiendo una fácil interpretación (Booth, Maindonald, & Smith, 2002). Esto ha llevado a que el modelo de Lee Carter haya sido aceptado ampliamente en las Ciencias Actariales, llevando a su utilización en estudios referidos a diferentes países y épocas temporales.

Sin embargo, también han surgido varias críticas al modelo, siendo una de las más destacables que el método se realiza bajo la hipótesis de que los parámetros relacionados con la edad son invariantes en el tiempo. Esto provocaría la posibilidad de que las pautas históricas no fuesen válidas para el futuro y que, por lo tanto, se pasen por alto los cambios estructurales (Lee, Ronald, 2000).

Ello ha motivado a diferentes autores a sugerir modificaciones para aumentar la capacidad predictiva y el ajuste del modelo. Entre ellas, se destacan brevemente las siguientes aportaciones:

- Booth et al. (2002) proponen ajustar el parámetro  $\kappa_t$  mediante un modelo de regresión de Poisson al número anual de fallecimientos en cada edad,  $D_{x,t}$ . Esto permite dar mayor peso a las edades en las que el número de fallecimientos es grande y proporciona una forma más sencilla de cálculo en comparación a la minimización propuesta por Lee Carter, ya que no requiere encontrar la raíz de una función no lineal.

El modelo de Poisson utilizado sigue la siguiente expresión:

$$\ln[D_{x,t}] = \ln[N_{x,t}] + \ln[m'_{x,t}] + \epsilon'_{x,t}$$

donde  $N_{x,t}$  representa la población con edad  $x$  en el tiempo  $t$ ,  $\ln[m'_{x,t}] = \alpha_x + \beta_x \kappa'_t$ , haciendo  $\kappa'_t$  referencia a  $\kappa_t$  bajo este ajuste, y siendo  $\epsilon'_{x,t}$  los residuos después del ajuste a  $\kappa_t$ .

- Renshaw y Haberman (2006) sugieren añadir un nuevo término al modelo de Lee Carter, el cual recoge el efecto cohorte en la mortalidad. Su expresión generalizada se da de la siguiente forma:

$$\ln(m_{x,t}) = \alpha_x + \beta_x^1 \kappa_t + \beta_x^2 \gamma_{t-x} + \epsilon_{x,t}$$

donde  $\gamma_{t-x}$  representa el nivel de mortalidad general para la cohorte nacida en el año  $t - x$  y  $\beta_x^2$  mide la sensibilidad a cambios en  $\gamma_{t-x}$  dada la edad  $x$ .

- Currie, Durban, y Eilers (2004) tratan de estimar la superficie de la mortalidad. Para ello, proponen la utilización de splines de regresión bidimensionales, específicamente basic splines con penalizaciones, usualmente conocidas como penalised splines. Su expresión básica se reduce a la siguiente:

$$\ln(m_{x,t}) = \sum_{i,j} \theta_{i,j} B_{ij}^{ay}(x, t)$$

## **CAPÍTULO 3. REDES NEURONALES**

### **3.1. Redes neuronales biológicas y redes neuronales artificiales**

El cerebro humano actúa de forma diferente a cualquier tecnología conocida, de una manera sorprendentemente compleja y no lineal, que, mediante la organización de sus unidades básicas denominadas neuronas, es capaz de procesar órdenes de forma más rápida que cualquiera ordenador podría llegar a realizar, ya sea el reconocimiento de una imagen o un patrón.

El objetivo principal de una red neuronal artificial (RNA) es utilizar el mismo procedimiento que el usado por el cerebro humano para resolver problemas, y precisamente son sobresalientes en aquellas cuestiones cuya obtención de soluciones es compleja mediante una programación tradicional. Esta técnica aprende por sí misma, en vez de precisar de una programación determinada, por ello un concepto a destacar es el de la experiencia.

Mediante la experiencia, la cual es desarrollada a lo largo del tiempo, el cerebro humano actúa de una forma determinada. La neuroplasticidad, o plasticidad cerebral, hace referencia a este concepto, y se define como la forma en la que el sistema nervioso se desarrolla o cambia a partir de su interacción con el entorno. Se podría afirmar que de la misma forma que la experiencia es esencial en el cerebro humano para el procesamiento de la información y el funcionamiento de sus neuronas, también lo es para una red neuronal artificial. Es por ello, la motivación hacia el trabajo en las redes neuronales artificiales (RNA), o comúnmente llamadas en la actualidad, redes neuronales.

Dada la relevancia que tiene la neurobiología en las redes neuronales artificiales, es obligado hacer referencia a D. Santiago Ramón y Cajal, eminente médico e investigador español y padre de la neurociencia. Sus estudios en este campo le valieron el premio Nobel de Medicina en el año 1906 y su trabajo permitió la comprensión de la estructura del sistema nervioso (Ramón, 1906). Desarrolló la denominada teoría de la “doctrina de la neurona”, en la que se defiende, a través del estudio de los procesos conectivos de las neuronas, que el tejido cerebral está compuesto por células individuales.

Las redes neuronales biológicas están formadas por un elevado número de células llamadas neuronas, cuya tarea principal es recibir, procesar y transmitir la información.



Analizando su morfología básica, una neurona está compuesta por un cuerpo celular o soma y una serie de prolongaciones cortas llamadas dendritas. Estas, actúan como zona de recepción o entrada de los impulsos hacia el cuerpo celular. La extensión larga o axón, permite la transmisión de los impulsos desde el soma hacia otra neurona.

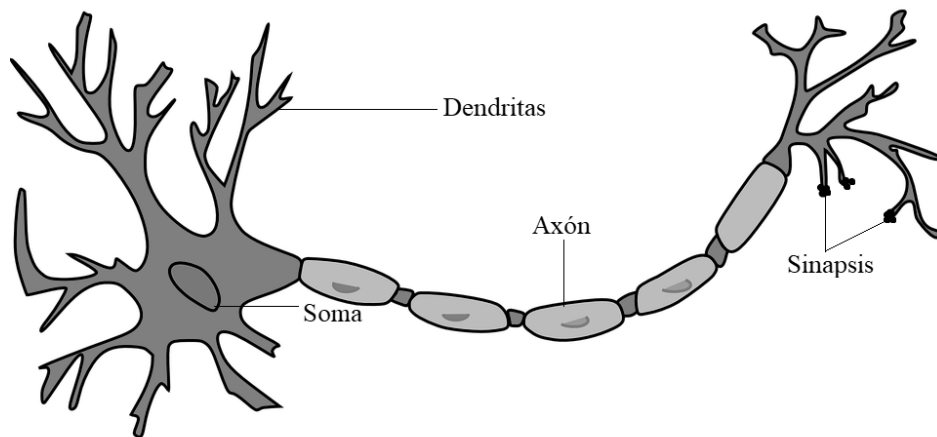


Figura 1. Morfología de una neurona humana.

Las neuronas se comunican entre sí mediante señales eléctricas cortas o picos de energía, y estos contactos interneuronales se realizan a través de la sinapsis, una unión electroquímica del axón de una neurona con las dendritas de otra. Cada neurona establece miles de conexiones con otras neuronas, y por ello, recibe constantemente señales entrantes que llegan al cuerpo de la célula. Podemos encontrar tres componentes principales en la sinapsis:

- Neurona presináptica: neurona que envía la información.
- Espacio sináptico: espacio existente entre dos neuronas. En él se produce la transmisión de información.
- Neurona postsináptica: neurona que recibe la información.

En una sinapsis, la neurona presináptica libera un neurotransmisor que llegará a los receptores de la membrana postsináptica, y que efectuará cambios en el potencial de membrana de la neurona receptora. Si la suma espacial y temporal de los cambios electroquímicos, también llamados señal, supera un determinado umbral, entonces, se generará un impulso de respuesta; este será transmitido a través del axón. La generación del impulso depende bien de la producción de un efecto inhibitorio (impidiendo la propagación de la información hacia la neurona postsináptica), o bien, la generación de un efecto excitatorio (creación del impulso de transmisión de la información). Dicho

tipo de procesamiento y arquitectura es el que se utiliza también en las redes neuronales artificiales, y debido a la importancia que reside en las conexiones interneuronales, a este tipo de estudios se le conocen como conectivismo.

El entendimiento sobre las redes neuronales biológicas permite la creación de una red neuronal artificial, siguiendo por ello los siguientes principios (Rebala, Ravi, & Churiwala, 2019):

- Se debe crear una serie de neuronas artificiales que estén interconectadas, y a su vez que estas interconexiones tengan unos pesos asignados.
- Debe existir un conjunto de neuronas de entrada, cuya función única será propiciar información a las neuronas que existen dentro de la red.
- También es preciso un conjunto de neuronas de salidas, las cuales transmiten información que proviene del interior de la red.
- La red se entrenará para alterar los pesos de las interconexiones de las neuronas, basándose en una concordancia input/output, también llamado aprendizaje.

### **3.2. Definiciones**

Existen numerosas definiciones acerca de las redes neuronales artificiales, pero podemos destacar principalmente las siguientes:

- “Una red neuronal artificial consiste en un conjunto de unidades de procesamiento (nodos) que simulan neuronas y están interconectadas mediante un conjunto de “pesos” (análogos a las conexiones sinápticas del sistema nervioso) de forma que permite que las señales viajen a través de la red neuronal.” (Fausett, 1994)
- “Una red neuronal es un conjunto interconectado de nodos de procesamiento simples, cuya funcionalidad está basada en las neuronas humanas. La capacidad de procesamiento de la red se almacena en los pesos, obtenidos por un proceso de aprendizaje de un conjunto de patrones de entrenamiento.” (Gurney, 1997)
- “Una red neuronal es un procesador distribuido masivamente en paralelo, compuesto por unidades de procesamiento simple, que tiene una propensión natural a almacenar el conocimiento experimental y a ponerlo a disposición para su uso.” (Haykin, 2010)

Por lo tanto, la semejanza con el cerebro viene dado por dos vertientes:

- En primer lugar, la red neuronal, mediante un proceso de aprendizaje adquiere conocimiento de su entorno.
- En segundo lugar, el conocimiento adquirido se almacena en unos denominados pesos sinápticos, y mediante un algoritmo de aprendizaje, se modifican esos pesos para llegar al objetivo deseado. Este método de ajuste de pesos se conoce como el más tradicional para el diseño de redes neuronales.

### **3.3. Propiedades**

Las redes neuronales artificiales tienen notables propiedades y beneficios, destacando primordialmente su gran capacidad computacional, ya que poseen una distribución en paralelo y una gran habilidad para aprender (Haykin, 2010). En detalle, las principales capacidades de las redes neuronales son:

- Generalización: las RNA son capaces de generar outputs razonables para inputs que no se han añadido en el proceso de aprendizaje o entrenamiento. Esta habilidad hace que puedan encontrar soluciones a problemas de gran escala.
- No linealidad: ésta se distribuye a lo largo de la red en el espacio, ya que las interconexiones de la red en sí misma no son lineales. La importancia de esta propiedad reside en que la red podrá procesar señales de entrada que no sean intrínsecamente lineales.
- Adaptabilidad: las RNA cambian sus pesos sinápticos según se vayan produciendo cambios en el entorno. Además, es posible reentrenar una red neuronal para detectar y hacer frente a pequeños cambios en su entorno general. Cuanto más adaptable sea la red, más fuerte será su performance cuando tenga que operar en un entorno no estacionario.
- Información contextual: cada neurona de la red se ve influenciada por la actividad de las otras neuronas constituyentes de la red neuronal, permitiendo que se trate el problema en un ámbito global.
- Tolerancia al error: debido a que una red neuronal distribuye la información almacenada en neuronas y uniones, cuando se produce un error o condición adversa, generalmente solo producirá una disminución del rendimiento más que un fallo de gran magnitud. Esto hará que la red neuronal tenga la interesante propiedad de trabajar incluso si los datos son incompletos o tienen ruido.

- Universalidad en el análisis y el diseño: en el sentido de que, en todas las aplicaciones de redes neuronales, encontraremos a las neuronas como el componente principal que hará posible compartir teorías, algoritmos y sus diferentes aplicaciones, permitiendo una fácil divulgación de la información.

### **3.4. Morfología de una neurona**

Una neurona es una unidad de procesamiento de la información que desempeña un trabajo fundamental en la aplicación de una red neuronal, es por ello una unidad básica. Cada unidad realiza un trabajo concreto, el cuál es recibir un input de una fuente (ya sea un recurso externo u otra neurona), y usarlo para formar una señal de output que será transmitida a otra unidad. El sistema se considera paralelo debido a que cada neurona realiza sus propias funciones a la vez que el resto de neuronas que se contienen dentro de la red.

Podemos distinguir tres tipos de neuronas (Kröse, van der Smagt, & Smagt, 1993):

- Neuronas de entrada, cuya función principal será recibir información desde el exterior de la red neuronal y transmitirla al interior.
- Neuronas ocultas, cuyos inputs y outputs permanecen dentro de la red neuronal.
- Neuronas de salida, encargadas de enviar información desde el interior hacia el exterior de la red neuronal, generando unas predicciones en forma de valor real o decisión booleana.

El esquema mostrado a continuación, presenta el modelo interno de una neurona, y podemos destacar los siguientes elementos básicos.

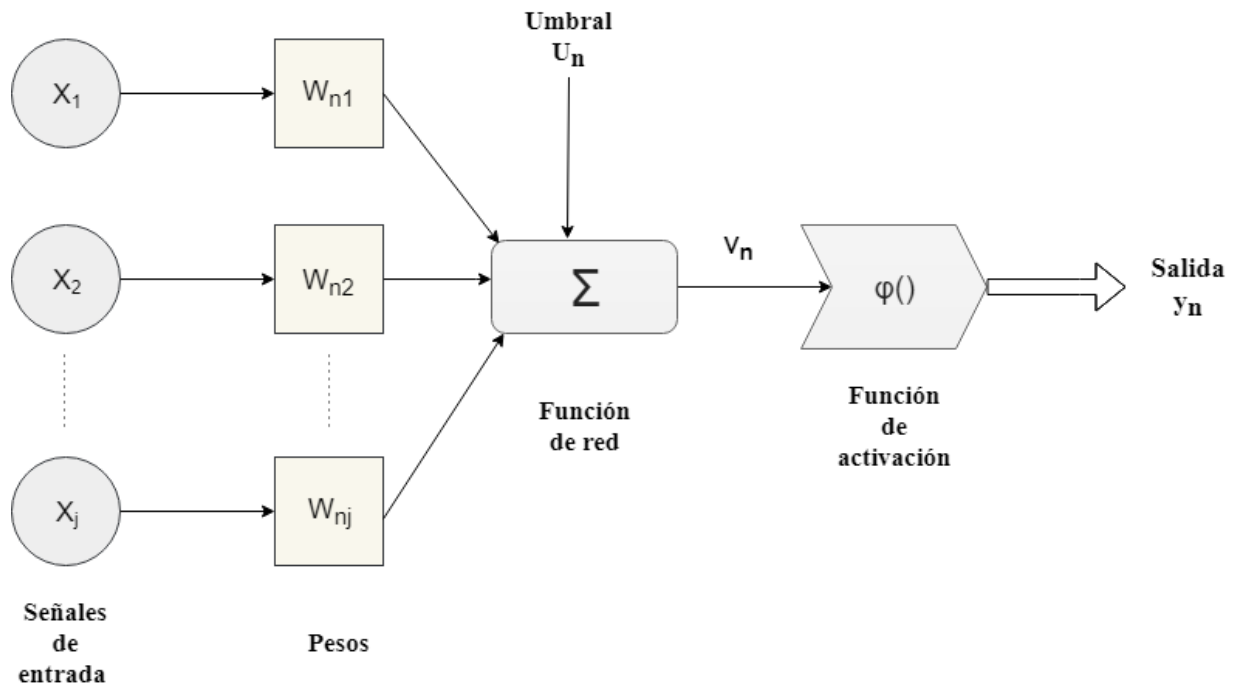


Figura 2. Esquema básico de una neurona.

- Conexiones entre neuronas, las cuales son medidas con pesos. Cada señal de input  $x_1, x_2 \dots x_j$  conectada con la neurona  $n$  será multiplicada por un peso  $w_{n1}, w_{n2} \dots w_{nj}$ .
- Una función de base o red, que será la encargada de proporcionar una salida de procesamiento a los inputs de la neurona  $n$ .  $S_n$  es, por lo tanto, una combinación lineal dadas las señales de entrada y los pesos, también llamado potencial de activación.

$$S_n = \sum_{j=1}^m w_{nj} x_j$$

- El potencial de activación  $S_n$  será sumado a un umbral o parámetro externo, denotado por  $U_n$ , que se encargará de realizar una transformación afín. Dependiendo de si tiene un valor positivo o negativo, estaremos hablando de un efecto excitatorio o inhibitorio correspondientemente. Normalmente, el valor el umbral es inicializado en 1.

$$V_n = S_n + U_n$$

- Una función de activación  $\varphi$ , la cual se encargará de tomar el total de output  $V_n$ , y transformarlo en un nuevo valor, de forma que la amplitud del rango final se vea limitado como un valor finito.

$$y_n = \varphi(U_n + S_n)$$

$$y_n = \varphi(V_n)$$

### 3.5. Tipos de funciones de activación

Las funciones de activación son un eje central en las estructuras de redes neuronales. Sin ellas todo el procedimiento que realiza una red neuronal se asemejaría a utilizar funciones lineales, pero sabemos que la gran mayoría de problemas que intentan resolver son no lineales y de una naturaleza compleja. Podemos distinguir diferentes tipos, de los que podemos destacar:

Funciones binarias: es una función de activación basada en un umbral. Si el valor de entrada es no negativo, el valor de salida adquirirá el valor de 1. En caso contrario asumirá un valor de 0. Por ello su rango está definido en el intervalo (0,1). Un inconveniente de este tipo de funciones es que solo permite valores booleanos como output, limitando su uso a esquemas binarios, cómo podría ser un problema de clasificación en dos grupos (Ciaburro, 2017).

$$y_n = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Función lineal: es una de las funciones de activación más sencillas. Simplemente multiplica la información de entrada por los pesos de la neurona, creando un impulso proporcional al input. El output estará definido al igual que el input en el rango  $(-\infty, +\infty)$ . El mayor inconveniente de este tipo de función es que no permite la utilización del algoritmo de retro propagación al no ser diferenciable, y por ello, al derivar parcialmente lo que se obtiene es una constante la cual no encuentra relación con la información de entrada (MissingLink, 2018).

$$\varphi(x) = ax$$

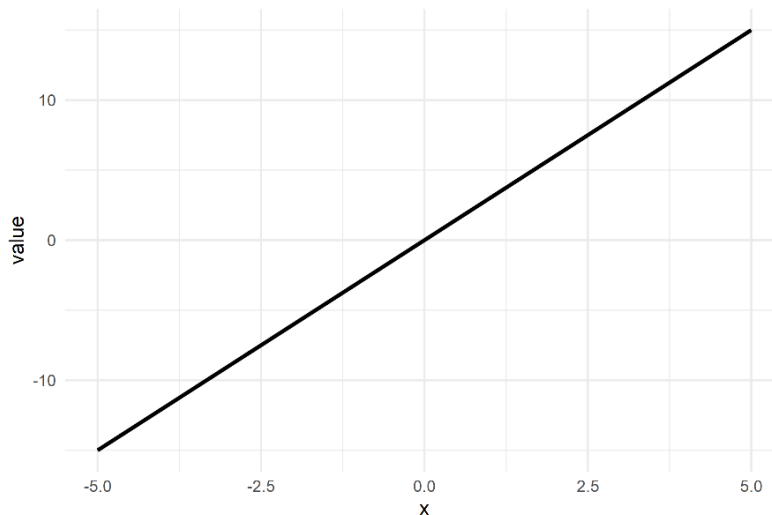


Figura 3. Representación de una función lineal.

Funciones no lineales: la mayoría de redes neuronales actuales utilizan este tipo de función de activación. Esto es debido a que son diferenciables y por lo tanto permiten el uso del algoritmo de retro propagación, ya que va a ser posible obtener una derivada que conecte con el input. Podemos hacer la siguiente clasificación:

- Función sigmoide: es probablemente la función de activación más conocida. Como principales características podemos encontrar que es continua y estrictamente positiva, con una forma típica de S. Su derivada existe en todos los puntos, y por lo tanto es diferenciable. Podemos diferenciar dos subtipos:
  - o Logística: está definida en un rango entre 0 y 1, teniendo la capacidad de normalizar o “aplstar” la información que se obtendrá como salida. Como inconvenientes, presenta que el output no está centrado en el 0, y el conocido problema de desvanecimiento de gradiente en computación. Además, es bastante lenta a la hora de procesar la información y converger.

$$\varphi(x) = \frac{1}{1 + \exp(-ax)}$$

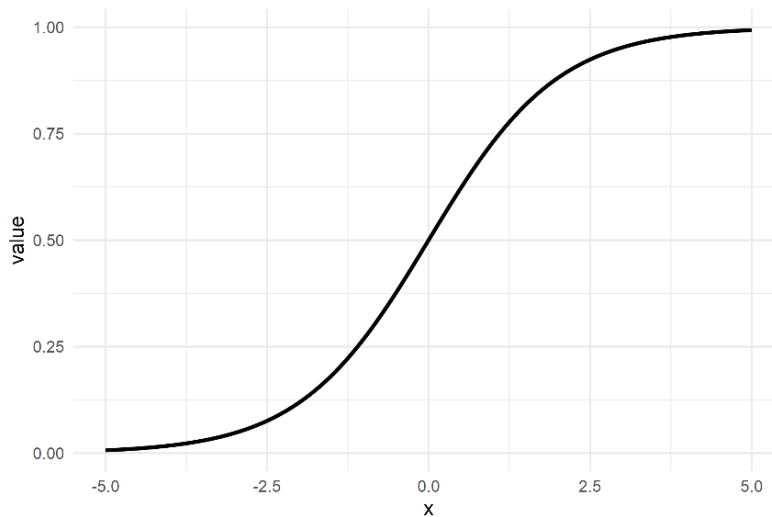


Figura 4. Representación de una función logística.

- Hiperbólica tangente: este tipo de función es muy parecida en todos los ámbitos a la logística, asemejándose a una versión escalada de ella, pero se diferencia en que está centrada en el cero y su rango se define de -1 a 1.

$$\varphi(x) = \tanh(x)$$

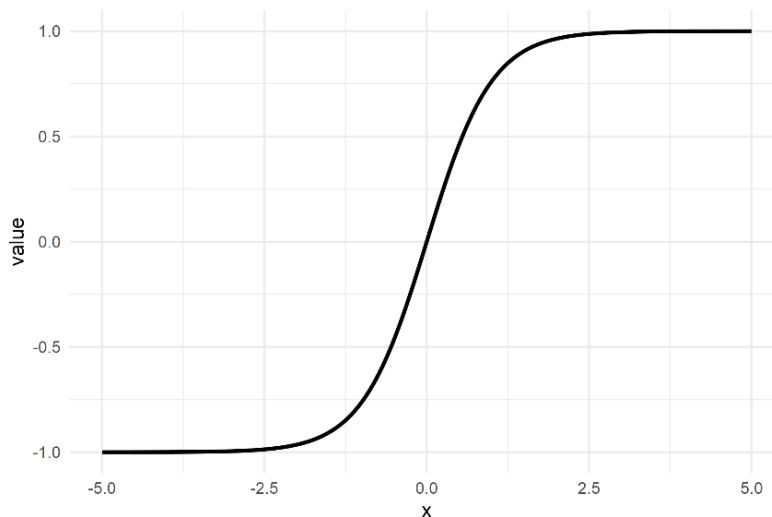


Figura 5. Representación de una función hiperbólica tangente.

- ReLU (Rectified Linear Unit) y sus variantes: es una función continua con rango definido en  $(0, +\infty)$ . Ha obtenido gran popularidad en los últimos años debido a que en la práctica es simple de aplicar y más rápida de procesar que las



sigmoides, siendo la función de activación preferida en la actualidad para problemas de Deep Learning (Ciaburro, 2017).

$$\varphi(x) = \max(0, x)$$

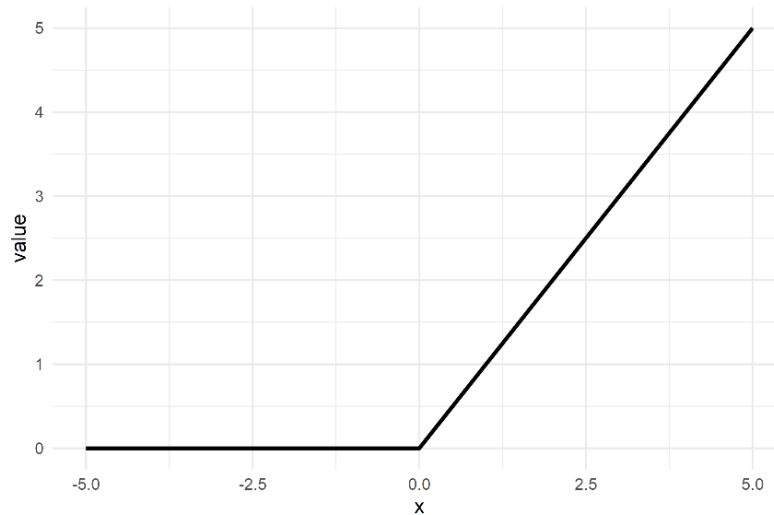


Figura 6. Representación de una función ReLU.

### 3.6. Topología de las redes neuronales

En los párrafos anteriores, se hizo una profundización en el concepto de neurona o unidad de procesamiento. El patrón con el cual las neuronas establecen sus conexiones y su estructura para formar redes, hará que podamos distinguir diferentes topologías, de las cuales destacaremos (Haykin, 2010):

- Red neuronal de avance, las cuales se caracterizan en que la información se propaga de las neuronas de entrada a las neuronas de salida estrictamente “hacia delante”. La estructura más simple de este tipo de red es aquella que tiene una capa de entrada, la cual solo recibe información que se propaga directamente a una capa de salida, encargada de procesar (Kröse et al., 1993). Se puede distinguir dos subtipos de estructuras dependiendo de la distribución en forma de capas que adquieran las neuronas.
  - o Red neuronal de una sola capa: recibe el nombre debido a que solo posee una capa de salida, la cual es encargada de realizar los cálculos. De forma general, la capa de entrada no se contabiliza en las estructuras ya que no

se realiza ninguna computación, simplemente transmite la información sin alteración.

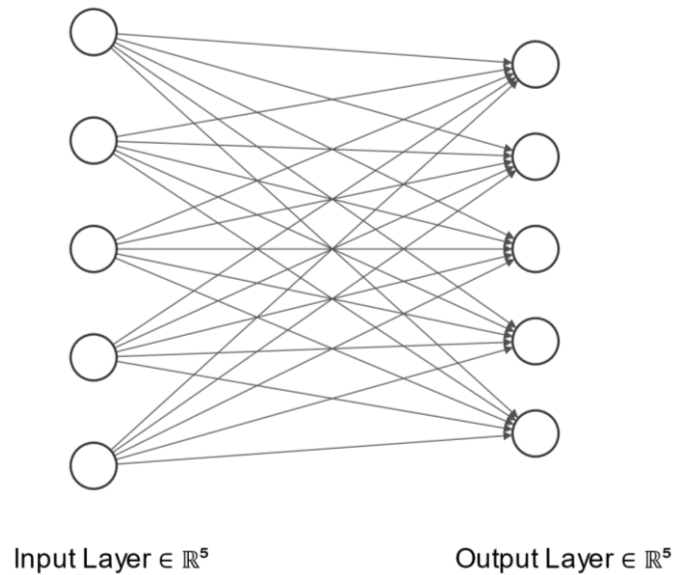


Figura 7. Diagrama de una red neuronal de una sola capa.

- Red neuronal multicapa: este tipo de topología se da cuando existe una o más capas ocultas. El concepto de “oculto”, hace referencia a que es aquella parte de la red que no puede observarse ni en forma de input ni como un output. Si la red solo tiene una capa oculta, se le conoce como Vainilla, y su estructura básica sería una capa de entrada, que recibe la información que actuará como input, una capa oculta, y una capa de salida que generará la información predicha. Si la red tiene dos o más capas ocultas, se le conoce como una Red Neuronal Profunda. La red es más precisa cuanto mayor sea el número de capas, siendo un número óptimo entre 3 y 10, ya que a partir de la decena la eficacia predictiva empieza a disminuir (Goodfellow, Bengio, & Courville, 2016).

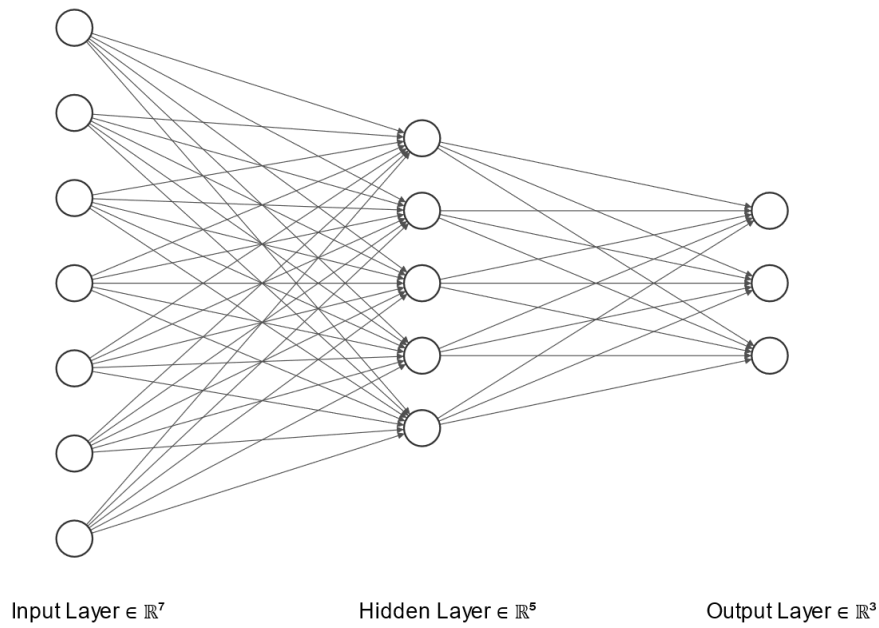


Figura 8. Diagrama de una red neuronal con una capa oculta.

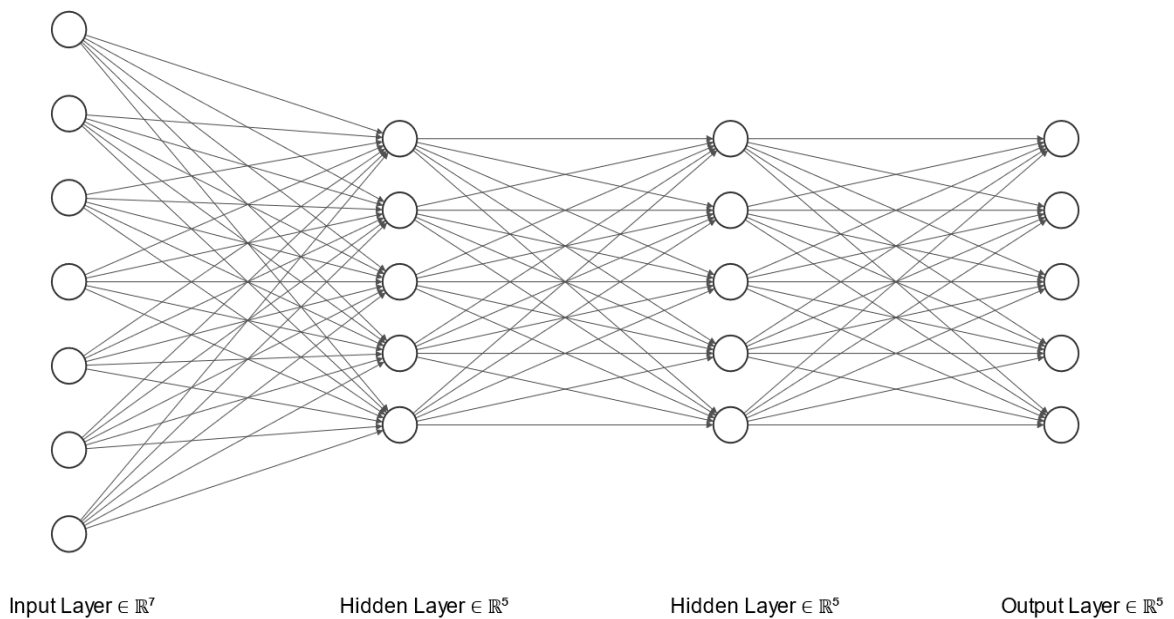


Figura 9. Diagrama de una red neuronal con dos capas ocultas.

- Red neuronal recurrente: la cual contiene una retroalimentación en sus conexiones, comportándose de una forma dinámica. La estructura más simple es en la cual una sola capa de neuronas emite outputs, que posteriormente serán los inputs, y así sucesivamente. En cada bucle de retroalimentación, hay unas unidades de retardo en el tiempo, que dan como resultado un comportamiento

dinámico sin linealidad. En apartados posteriores se podrá observar en detalle la representación gráfica de esta topología de RNA.

### **3.7. Entrenamiento de una red neuronal**

Las redes neuronales deben de ser programadas para que, dado un cierto conjunto de entradas de información, realice la salida u output que se desee. Para ello es necesario que aprenda de su entorno y al igual que ocurre con el cerebro humano, existen diferentes formas de adquirir este conocimiento.

#### Aprendizaje supervisado

El aprendizaje supervisado implica que los pesos sinápticos se vayan modificando mediante el uso de ejemplos de entrenamiento, los cuales consistirán en un conjunto de datos que se asocian a unas respuestas correctas. La red se entrena tantas veces sea necesario para que se alcance un estado en el que ya no haya cambios significativos en los pesos, y, por lo tanto, sean óptimos. Una vez los pesos sean los óptimos, es cuando se le permite a la red que pueda seguir trabajando en el entorno, pero sin supervisión. Así pues, en la siguiente ocasión que se le suministre al algoritmo un conjunto de datos nuevo, este debería de ser capaz de predecir la respuesta correcta en base a las características de las que ha aprendido anteriormente.

Podemos dividir las cuestiones tratadas en este método en dos categorías (Rebala et al., 2019):

- Problemas de clasificación, los cuales se basan en dada una serie de categorías, intentar clasificar algo dentro de ellas. Un ejemplo en computación sería cuando una red neuronal es capaz de etiquetar correctamente una imagen.
- Problemas de regresión, encargados de predecir variables continuas. La predicción del precio de un índice bursátil o del número de transacciones realizadas puede ser un ejemplo de ello.

#### Aprendizaje no supervisado

Al contrario del aprendizaje supervisado, no hay un set a priori de categorías o ejemplos que la red pueda utilizar para optimizar sus pesos. Dada una gran cantidad de datos, el

algoritmo intentará encontrar similitudes y a partir de ello crear grupos de clasificación o clústeres.

Un ejemplo de aprendizaje no supervisado podría ser cuando se le proporciona a una red neuronal una gran batería de imágenes de perros y gatos sin etiquetar. Después de que el algoritmo identifica características en estas imágenes y encuentre similitudes, si le proporcionamos una imagen de un perro será capaz de clasificarla en su grupo correspondiente, pero como tal no sabrá si es un perro o un gato, solo que comparte ciertas características iguales a un grupo concreto de imágenes.

### Aprendizaje semi supervisado.

Este tipo de método de aprendizaje utiliza los dos mencionados anteriormente, y por ello se le dará a la máquina un gran conjunto de datos de los que solo algunos estarán etiquetados. El propósito de esto es que el algoritmo sea capaz de crear clústeres, y utilizar los pocos datos proporcionados identificados para asignar dichas etiquetas a los grupos creados. Entre sus beneficios, se encuentra que este método puede ser muy útil cuando poseemos un amplio conjunto de datos en el que solo algunos de ellos están etiquetados debido a una escasa calidad de ellos. Además, permite ahorrar tiempo si es requerido manualmente identificar todos y cada uno de los datos.

### Aprendizaje por refuerzo

En este tipo de aprendizaje, la asociación de señales de entrada con las de salida viene dada mediante una continua interacción con el entorno (Haykin, 2010). Se utiliza en casos que impliquen situaciones de cambio constante, como por ejemplo podría ser el ajedrez. En él, nos encontramos que hay infinitas posibles respuestas que son imposibles de enumerar, ya que el entorno externo cambia constantemente y es impredecible. Es por ello que se hace preciso el uso de un algoritmo de aprendizaje que detecte la situación externa y elija en base a ello una acción acorde a su propia experiencia. Cuanto más tiempo disponga para aprender, mejor podrá valorar el impacto a largo plazo de su decisión y más precisa será (Rebala et al., 2019).

### 3.8. Algoritmo de retro propagación

El algoritmo de retro propagación del error (McClelland, Rumelhart, & PDP Research Group, 1986) es el método de entrenamiento más popular en RNA. Este utiliza aprendizaje supervisado, lo que significa que se le proporciona un conjunto de ejemplos de entrenamiento hasta que la red neuronal aprenda. Podemos distinguir dos fases al utilizar este algoritmo:

- En primer lugar, existe una etapa de funcionamiento, en donde se suministra la información de entrada, inicializada con unos pesos y umbrales aleatorios, la cual se verá procesada a través de las capas ocultas, y finalmente mediante la capa de salida se obtendrá un output determinado. Este output se comparará con el output objetivo, el cual estará predefinido dado un cierto conjunto de inputs, obteniendo un error basado en la diferencia.
- Aquí entra en acción la etapa de aprendizaje o entrenamiento, donde el error calculado se propaga hacia atrás mediante una señal de error, desde la capa de salida a la capa oculta, y de esta, a la capa de entrada. La derivada parcial de la función de activación es utilizada para calcular la señal de error, y mientras va retrocediendo cambia los pesos sinápticos que permanecían almacenados en la estructura de la red. Este ciclo compuesto de ir hacia delante y luego hacia atrás se llama epoch o ciclo de entrenamiento. La red irá realizando dichos ciclos hasta que el error esté minimizado, y entonces se podrá decir que la red está entrenada, almacenando los pesos óptimos.

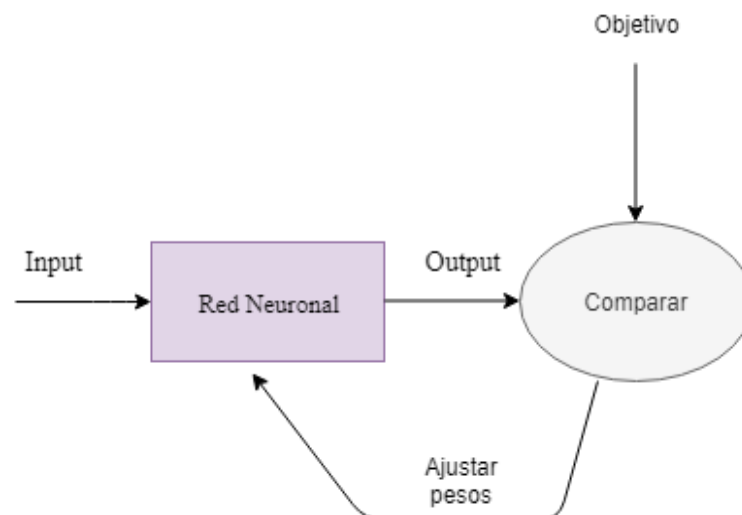


Figura 10. Esquema sobre el entrenamiento de una red neuronal.

En términos matemáticos, la etapa de funcionamiento es similar a la de una sola neurona descrito en apartados anteriores, pero en este caso lo extenderemos a una estructura multicapa vainilla:

Dado un conjunto  $p$  de datos de entrada  $X^p: x_1^p, x_2^p \dots x_i^p$ , serán transmitidos desde las neuronas de la capa de entrada  $i$  hacia las de la capa oculta  $j$  mediante los pesos  $w_{ji}$ . El total de información  $\theta_j^p$  que recibe una neurona de la capa oculta  $j$  es:

$$\theta_j^p = \sum_{i=1}^N w_{ji} x_i^p + U_j$$

donde  $U_j$  es el umbral, el cual actúa como un peso adicional externo y se le asocia un valor inicial de 1.

En la capa oculta se realizará la transformación de las señales mediante la función de activación  $\varphi(\cdot)$ , que transmitirá el valor de salida  $y_j^p$ .

$$y_j^p = \varphi(\theta_j^p)$$

El valor de salida  $y_j^p$ , lo recibirá una neurona  $k$  de la capa de salida a través de unos pesos  $v_{kj}$  y un umbral  $U_k$ .

$$\theta_k^p = \sum_{j=1}^H v_{kj} y_j^p + U_k$$

Por último, a través de otra función de activación, se obtendrá el output final o valor de salida de la neurona de salida  $k$ ,  $y_k^p$ .

$$y_k^p = \varphi(\theta_k^p)$$

En la etapa de aprendizaje, el objetivo es hacer mínima la diferencia entre el output objetivo del ejemplo de entrenamiento suministrado y el obtenido por la red. Por ello, el entrenamiento de una red neuronal se basa en un problema de optimización sin restricciones. Dada la función de error:

$$E^p = \frac{\sum (g_k^p - y_k^p)^2}{2}$$

donde  $g_k^p$  es el valor objetivo o deseado del conjunto de entrenamiento  $p$ , podemos afirmar que  $E^p$  está en función de todos los pesos de la red. Además, podemos definir:

$$E = \sum_p E^p$$

como la suma del error cuadrático.

El fin de este algoritmo es encontrar los pesos sinápticos que minimicen el valor de  $E^p$  y el método utilizado para ello se conoce como descenso de gradiente.

Un gradiente es el conjunto de todas las derivadas parciales de una función, y se puede definir básicamente como la pendiente de una curva. En el caso del algoritmo de retro propagación, el interés recae en el gradiente de la función de coste o función de error  $E$ . El objetivo de realizar el descenso de gradiente en cada paso es encontrar el mínimo global de la función de coste, donde el error es más bajo. Justo en ese punto es donde el modelo se ajusta de la mejor forma posible a los datos, y por ende las predicciones serán más acertadas. Cuando el gradiente toma una dirección negativa, se estará determinando el decremento más rápido en el error. Por lo tanto, la idea está en que el gradiente descendente de  $E^p$  sea un vector igual a la derivada parcial negativa de  $E^p$  con respecto a cada uno de los pesos.

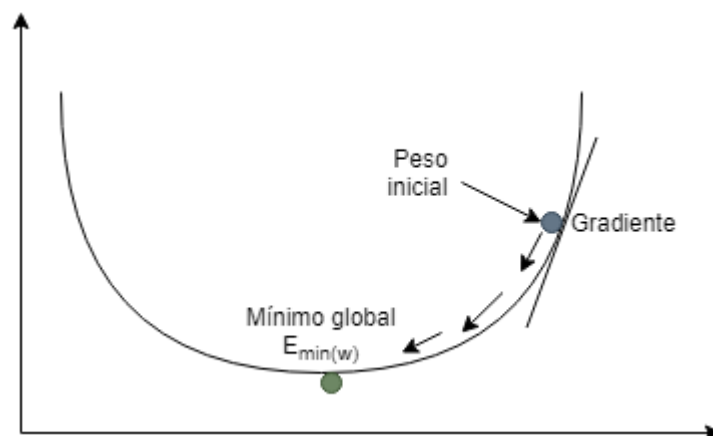


Figura 11. Representación gráfica del descenso del gradiente.

La corrección de los pesos en una neurona de la capa de salida se realizaría mediante la regla delta:



$$\Delta v_{kj} = -\eta \frac{\partial E^p}{\partial v_{kj}} = \eta \sum_{p=1}^P \delta_k^p y_j^p$$

donde

$$\delta_k^p = (g_k^p - y_k^p) \varphi'(\theta_k^p)$$

y  $\eta$  es la tasa de entrenamiento o constante de proporcionalidad.

En una neurona de la capa oculta, se realizaría el procedimiento utilizando la regla de la cadena.

$$\Delta w_{ji} = \eta \sum_{p=1}^P \delta_j^p y_i^p$$

donde

$$\delta_j^p = \varphi'(\theta_j^p) \sum_{k=1}^M \delta_k^p v_{kj}$$

Este procedimiento constituye la regla delta generalizada para una red neuronal compuesta de nodos no lineares.

Finalmente, cabe mencionar que, para la corrección de los pesos, podemos encontrar dos modalidades de aprendizaje:

- Aprendizaje por lotes o modo batch, el cual ajustará los pesos sinápticos después de que hayan sido revisados todos los ejemplos que constituyen un ciclo de entrenamiento.
- Aprendizaje on-line, en el cual los pesos se ajustan ejemplo a ejemplo, los cuales se presentarán en orden aleatorio.

## CAPÍTULO 4. REDES NEURONALES RECURRENTE

En las redes de propagación hacia delante podemos encontrar que su estructura unidireccional es fuertemente limitante en el desempeño de ciertas tareas. No obstante, crear redes en donde el procesamiento de una neurona afecta al de otra es posible.

Las redes neuronales recurrentes son una topología de RNA en las que las conexiones entre neuronas son cíclicas, es decir, existe un bucle de retroalimentación vinculado a sus decisiones pasadas, creando una secuencia temporal que denota un comportamiento dinámico. Están basadas en el trabajo de David Rumelhart en 1986 (Rumelhart, Hinton, & Williams, 1986), en el cual se muestra que las redes neuronales multicapas pueden aprender representaciones internas de datos, usando este enfoque para posteriores investigaciones sobre la cognición, como es el caso de la memoria.

A la hora del aprendizaje, no solo utiliza la información de entrada que se le suministra a la red, sino que también utiliza todo lo que ha experimentado con el paso del tiempo, haciendo de gran importancia la decisión que toma en un momento concreto ya que condiciona a la que tome inmediatamente después. Por ello, se puede decir que este tipo de red tiene memoria, tal y como la del cerebro humano.

Gracias a esta característica, las redes neuronales recurrentes pueden realizar ciertas funciones que una red de propagación hacia delante no podría, concretamente el reconocimiento de patrones como una secuencia de datos o modelización de secuencias. Algunas de las utilidades principales de esta topología de RNA son:

- Predicciones de series temporales
- Reconocimiento y predicción del habla
- Predicción del tiempo
- Reconocimiento y predicción de la escritura
- Secuenciación robótica
- Traducción del lenguaje
- Procesamiento de imágenes

Gráficamente, se puede observar en la Figura 12 la comparación entre una red neuronal artificial de propagación hacia delante y una RNR, radicando la diferencia en que las RNR utilizan el input suministrado, y además el output calculado previamente:

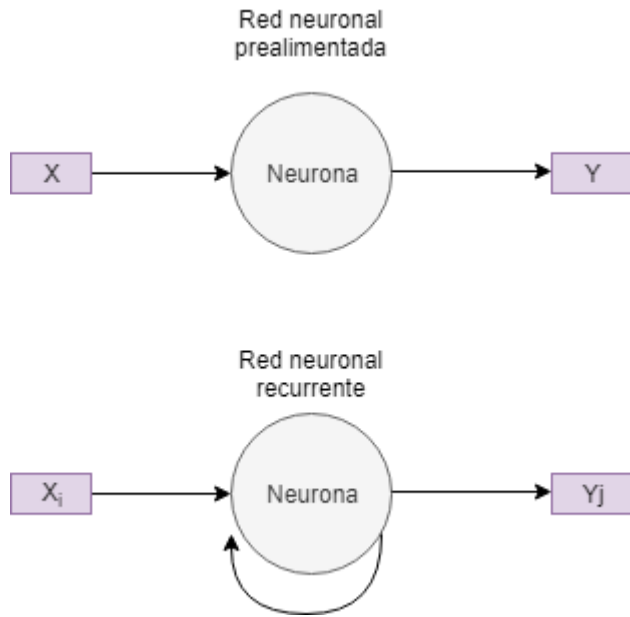


Figura 12. Comparación de una red neuronal básica con una recurrente.

La arquitectura típica de una red neuronal recurrente consiste en una red de redes neuronales, con inputs secuenciales dados a cada red además de la información procesada de la anterior secuencia (Rebala et al., 2019). Para entender esta naturaleza cíclica, se puede decir que una RNR simple está desplegada de la siguiente forma:

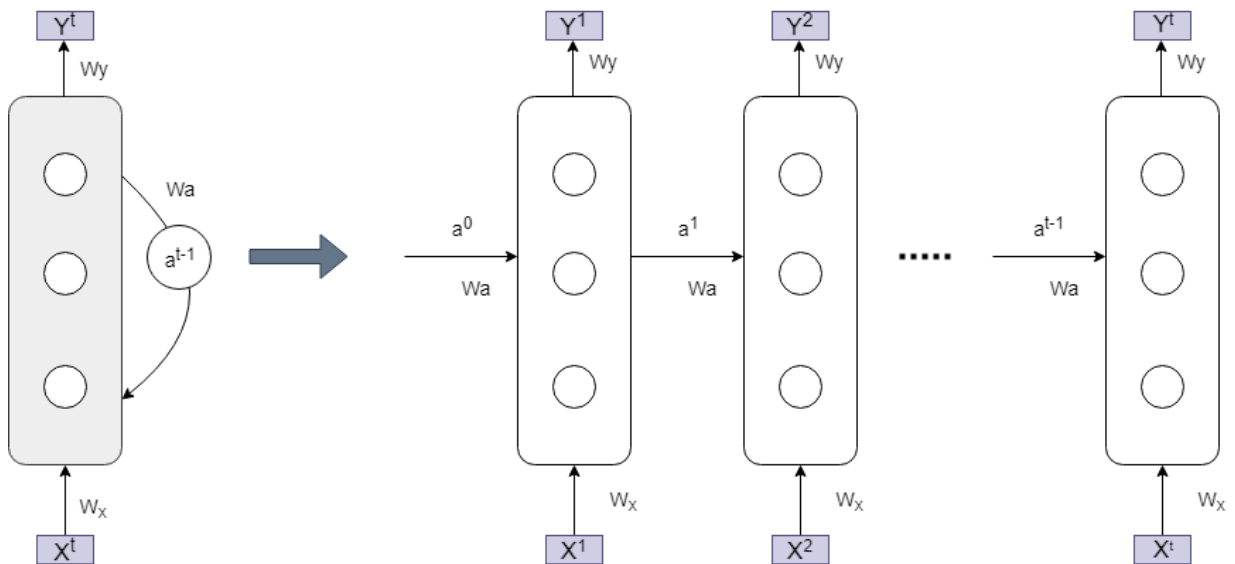


Figura 13. Despliegue de una red neuronal recurrente.

Se puede observar que se evalúa el estado de la red en diferentes periodos de tiempo (0,1,2...t), y por ello se dice que una red recurrente se conforma de varias copias de una red neuronal simple. La red neuronal recurrente calcula su estado con cada entrada de

la secuencia temporal, y pasa este estado al cálculo de la siguiente salida. De forma matemática, se puede expresar de la siguiente forma:

$$a^t = \varphi(W_a a^{t-1} + W_x X^t + U_a)$$

$$Y^t = \varphi(W_y a^t + U_y)$$

donde  $X^t$  es el vector de entradas en el periodo t,  $Y^t$  es el vector de salidas calculado en el periodo t,  $a^t$  es el input de activación calculado de la secuencia t,  $W_x$  es el vector de pesos para los inputs,  $W_y$  es el vector de pesos para el output,  $W_a$  es el vector de pesos para el input de activación y  $\varphi$  es la función de activación elegida. Además, encontramos los respectivos umbrales  $U_a$  y  $U_y$ .

La forma de aprendizaje de una red neuronal recurrente es similar al algoritmo de retro propagación visto anteriormente, solo que esta vez la corrección del término de error se realizará a través del tiempo. Por ello, el algoritmo recibe el nombre de retro propagación a través del tiempo.

Su secuencia matemática es similar al algoritmo de retro propagación tradicional, por lo que seguiría el siguiente pseudocódigo:

- En primer lugar, se descompone la red neuronal recurrente en n redes de propagación hacia delante.
- Se le asigna a la primera red un valor inicial aleatorio a los pesos.
- Seguidamente, se repite este proceso en cada red n desplegada.
- Llegado este momento, se introducen los inputs al conjunto de la red y se propagan de forma conectada, pasando por cada red n desplegada y calculando sus respectivos pesos.
- Se retro propaga el error hacia atrás en la red totalmente desplegada, y a cada paso por la red n se van actualizando los pesos al más óptimo.
- Finalmente, se realiza una media de todos los pesos finales óptimos de la red totalmente desplegada, para obtener el conjunto de pesos final de la red neuronal recurrente.

Dentro de las redes neuronales recurrentes, podemos encontrar diferentes tipos como las redes totalmente recurrentes, Elman y Jordan, Bidireccionales, Hopfield, Gated Recurrent Unit y Long Short-Term Memory. En este trabajo, nos centraremos en la última mencionada.

## CAPÍTULO 5. LONG SHORT-TERM MEMORY

En 1997, los investigadores alemanes Hochreiter y Schmidhuber publicaron el primer artículo que introduce las redes de memoria a corto-largo plazo, u originalmente, Long-Short-Term Memory (LSTM). Este tipo de red neuronal recurrente está estructurada de forma que es capaz de crear secuencias de datos y relacionarlas, permitiendo recordar o retener la información durante largos periodos de tiempo y reconocer patrones de comportamiento tanto a largo como a corto plazo. Dichos módulos de tareas son particularidades que una red neuronal recurrente tradicional encuentra dificultades a la hora de solventar.

Además, las redes LSTM son excelentes en resolver problemas de secuencias, en concreto:

- Clasificación de secuencias.
- Generación de secuencias.
- Predicción de secuencias.

Por lo que su aplicación en predicción de series temporales es eficaz. Además, esta tipología de RNA ha adquirido gran popularidad debido a que disminuye notablemente el conocido problema del desvanecimiento del gradiente en computación. En pocas palabras, la casuística concreta de este problema es que el gradiente se hace cada vez que se propaga a una capa más pequeño, provocando que el aprendizaje en las primeras capas sea extremadamente lento. Además, en redes neuronales recurrentes tradicionales este problema puede hacerse aún mayor ya que los gradientes se propagan hacia atrás a través del tiempo, no solo en capas. Gracias a la introducción de las neuronas con memoria a corto-largo plazo, se permite solucionar en gran medida la inestabilidad del gradiente.

En la Figura 14 se puede observar el interior de la arquitectura de una red LSTM. Lo más destacable de este tipo de red es la existencia de unas celdas cerradas encargadas de almacenar información, las cuales son capaces individualmente de administrar dichos datos mediante la apertura o el cierre de puertas. Además, permiten que posteriormente se puedan leer y recuperar la información almacenada para su uso. Esta característica de

almacenamiento a largo plazo es lo que diferencia las redes LSTM de las redes neuronales recurrentes tradicionales.

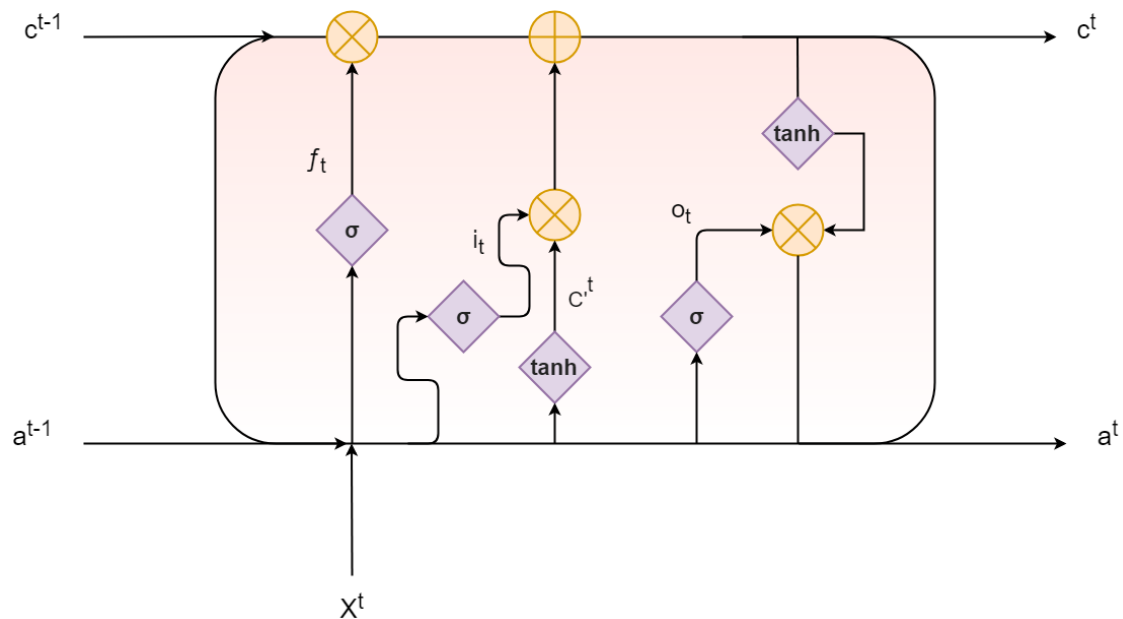


Figura 14. Esquema básico de una red LSTM.

Debido a su complejidad, es preciso ir explicando paso a paso su funcionamiento, así como los cálculos que se van realizando en paralelo.

El primer elemento que toma importancia es la llamada puerta de olvido o de actualización  $f_t$ , que se encuentra dentro la capa de olvido y se ve representada en la Figura 15. Su función es restablecer el valor del estado actual de la celda, es decir, determina en qué medida hay que olvidar los pesos que permanecen en ella. A través de una función  $\sigma$  sigmoide, se toma la decisión de olvidar o recordar para cada dato que se encuentra contenido en la celda, utilizando para estos cálculos  $X^t$  que es la entrada externa en la secuencia  $t$  y  $a^{t-1}$  que es la entrada de activación calculada en la secuencia anterior  $t-1$ , todo ello ponderado por sus respectivos vectores de pesos. El output de esta etapa queda acotado entre 0 y 1 debido a las características propias de la función de activación utilizada. Si la puerta da un resultado de 0, significa que la celda debería de olvidar todo lo que ha aprendido anteriormente, y si da un resultado de 1 tiene que recordarlo todo para su posterior uso.

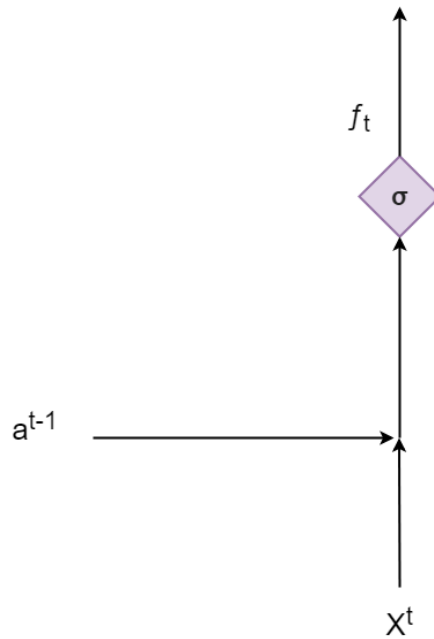


Figura 15. Representación de la capa de olvido en una red LSTM.

De forma matemática viene expresado como:

$$f_t = \sigma(W_f a^{t-1} + W_x X^t + U_f)$$

donde  $W_f$  son los pesos para calcular la puerta de olvido,  $W_x$  es el vector de pesos de las entradas y  $U_f$  es el umbral de la puerta de olvido.

El siguiente elemento y etapa relevante es la llamada puerta de entrada  $i_t$ , cuya función es seleccionar las entradas a la celda, es decir, determina qué información va a tener un impacto en la predicción que se va a realizar, y por ello la almacena en el estado actual de la celda. Esta se divide en dos pasos, tal y como se observa en la Figura 16.



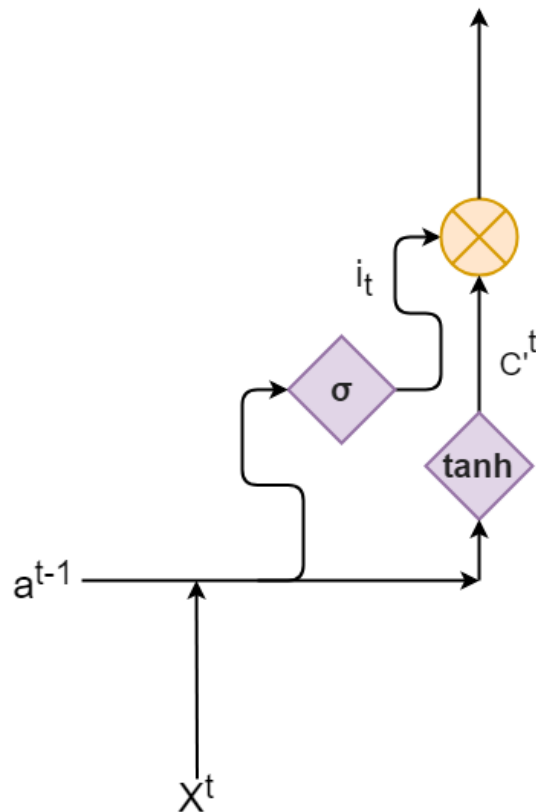


Figura 16. Representación de una puerta de entrada en una red LSTM.

- En primer lugar, mediante una función  $\sigma$  sigmoide se decide qué valores se van a actualizar de los contenidos en el estado actual celda. Para ello, se vuelve a hacer uso de  $X^t$  que es la entrada externa en la secuencia t y  $a^{t-1}$  que es la entrada de activación calculada en la secuencia anterior t-1 ponderado por sus respectivo vector de pesos.

$$i_t = \sigma(W_i a^{t-1} + W_x X^t + U_i)$$

donde  $W_i$  son los pesos para calcular la puerta de entrada y  $U_i$  es el umbral de la puerta de entrada.

- En segundo lugar, mediante una función tangente hiperbólica, se eligen nuevos valores candidatos  $c^t$  a entrar en el estado actual de la celda. El procedimiento de cálculo es similar al explicado anteriormente como se refleja en la fórmula:

$$c^t = \tanh(W_c a^{t-1} + W_x X^t + U_c)$$

donde  $W_c$  son los pesos para las entradas de activación y  $U_c$  es de nuevo el umbral.

A continuación, se realizan una serie de operaciones puntuales que cambiarán definitivamente el valor, ya antiguo, del estado de la celda  $c^{t-1}$  por el actual  $c^t$ , que será la salida de la celda en la secuencia t. Para ello, se utilizarán los valores obtenidos para la puerta de olvido, multiplicados por el estado anterior de la celda, eligiendo qué cantidad de información se va olvidar, para dar paso a los nuevos valores candidatos ponderados por la cantidad de información nueva que debe de entrar, tal y como se expresa en la siguiente fórmula:

$$c^t = f_t c^{t-1} + i_t c^t$$

De forma gráfica, podemos ver representado en la Figura 17 de color amarillo las operaciones puntuales de multiplicación y adición realizadas.

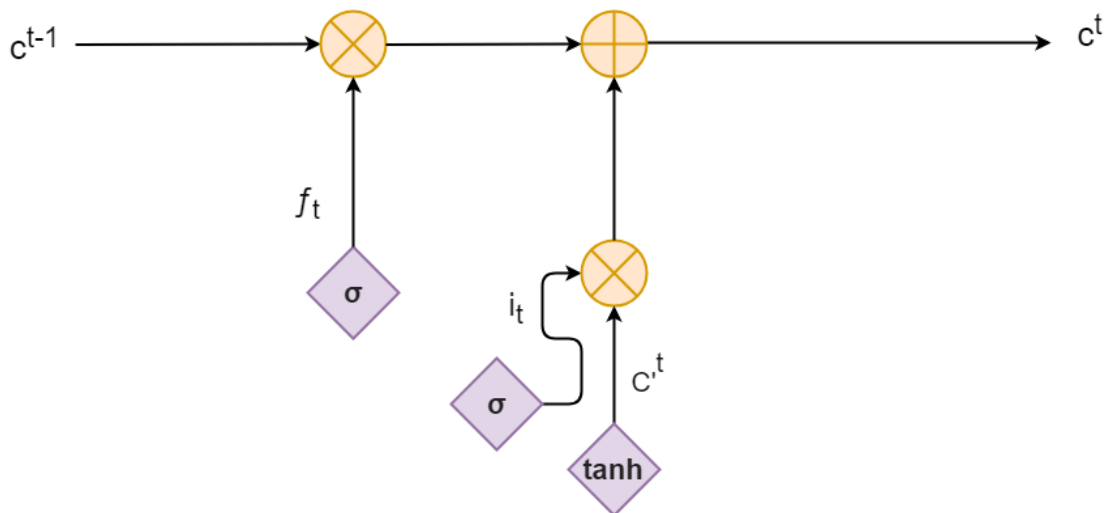


Figura 17. Representación de operaciones puntuales en una red LSTM.

Como paso final, se determina qué información tiene relevancia y debería ser un output en la capa de salida. Este proceso también se ve dividido en dos pasos como se observa en la Figura 18, y se explica a continuación:

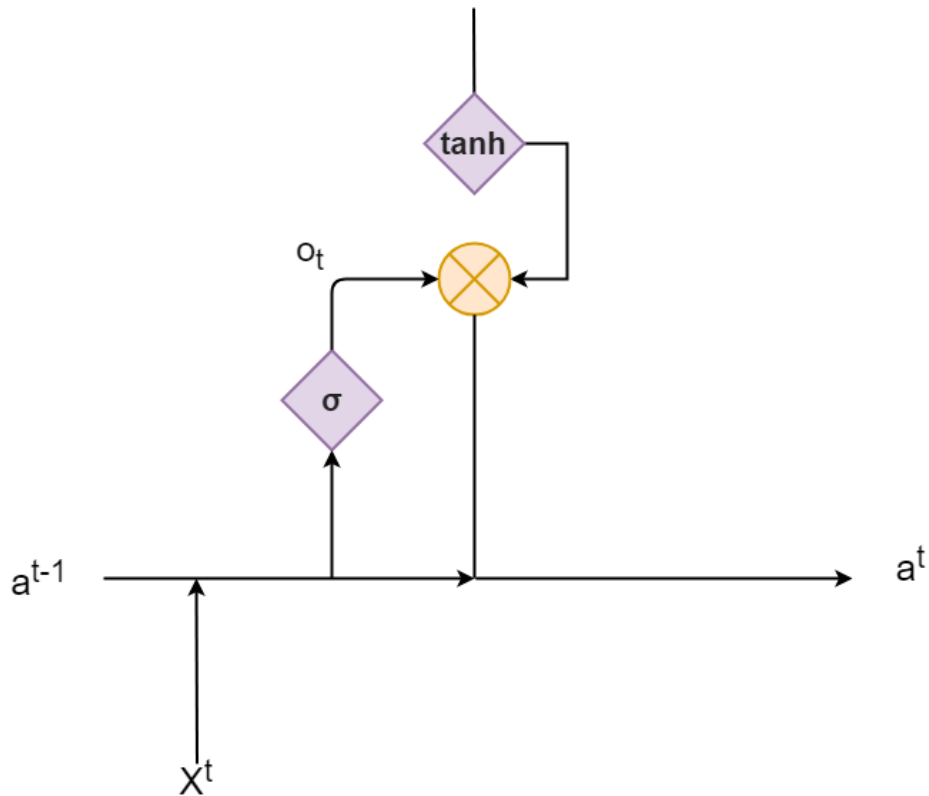


Figura 18. Representación de una capa de salida en una red LSTM.

- En primer lugar, mediante la puerta de salida, se va a determinar que parte de la información, de forma ponderada, puede ser un output, mediante el uso de:

$$o_t = \sigma(W_o a^{t-1} + W_x X^t + U_o)$$

donde  $W_o$  son los pesos para calcular la puerta de salida y  $U_o$  es el umbral de la puerta de salida.

- Seguidamente, a través de una función tangente hiperbólica, se acotan los valores del estado actual de la celda entre el rango típico de este tipo de función de activación, -1 y 1.

Finalmente, la versión depurada del output  $a^t$ , será la operación puntual de multiplicación entre los dos pasos mencionados anteriormente:

$$a^t = o_t \tanh(c^t)$$

y la salida final calculada en la secuencia  $t$  para la red neuronal al final de todo el proceso,  $Y^t$ , vendrá computada como:

$$Y^t = \varphi(W_y a^t + U_y)$$

donde,  $W_y$  el vector de pesos de las salidas y  $U_y$  su respectivo umbral.

## **CAPÍTULO 6. CASO PRÁCTICO: PREDICCIÓN DE LA MORTALIDAD.**

En el presente capítulo el objetivo es la predicción de la mortalidad aplicando, por un lado, el ya mencionado anteriormente modelo paramétrico de Lee Carter, y por otro, la red neuronal recurrente de tipo LSTM.

Para ello, se utilizarán los datos disponibles en Human Mortality Data Base para la población española, tanto femenina como masculina, entre los años 1950 y el último año disponible, que en este caso es el año 2016. Se han evitado los años anteriores a 1950 debido a que son periodos de distorsión en la demografía española a causa de la Guerra Civil.

Para evaluar el poder predictivo de ambas metodologías, se ha separado la muestra en dos partes:

- La primera, comprende el período desde 1950 hasta el año 2006, y serán los datos con los que se calibra los parámetros del modelo de Lee Carter o en su caso, se entrenará la red neuronal LSTM.
- La segunda, figura entre el año 2007 hasta el año 2016, y será aquella parte de la muestra que permite validar el poder predictivo de ambos métodos. En este caso, es conveniente analizar el error cuadrático medio en predicciones a 1 año, 5 años y 10 años.

Un primer paso importante es realizar un procesamiento de los datos para facilitar su posterior uso. Por ello, se ordena la muestra en un formato largo de cinco columnas, las cuales son género, año, edad, y un cálculo de la tasa de mortalidad y de la misma en términos logarítmicos. Además, se realiza la separación previamente mencionada entre el conjunto de entrenamiento o parametrización, y el conjunto de validación.

En la Tabla 1, se muestra un breve ejemplo del formato aplicado a los datos para el conjunto de entrenamiento. Cabe destacar, que la diferenciación presente en la columna “Género” viene dada ya que se ve a realizar el estudio de forma separada para hombres y mujeres. Una representación conjunta carecería de sentido ya que ambos géneros presentan tasas de supervivencia y mortalidad notablemente diferentes.

Tabla 1. Ejemplo del formato largo aplicado.

	<b>Genero</b>	<b>Año</b>	<b>Edad</b>	$m_{x,t}$	$\ln(m_{x,t})$
<b>1</b>	Femenino	1950	0	0.073338486	-26.126.698
<b>2</b>	Femenino	1950	1	0.013642755	-42.945.467
<b>3</b>	Femenino	1950	2	0.007390590	-49.075.477
<b>4</b>	Femenino	1950	3	0.004426429	-54.201.622
<b>5</b>	Femenino	1950	4	0.002995097	-58.107.788
<b>6</b>	Femenino	1950	5	0.002410560	-60.278.960

En la Figura 19, se observa una representación en forma de mapa de calor de la tasa de mortalidad logarítmica para el género femenino entre los años 1950 y 2016.

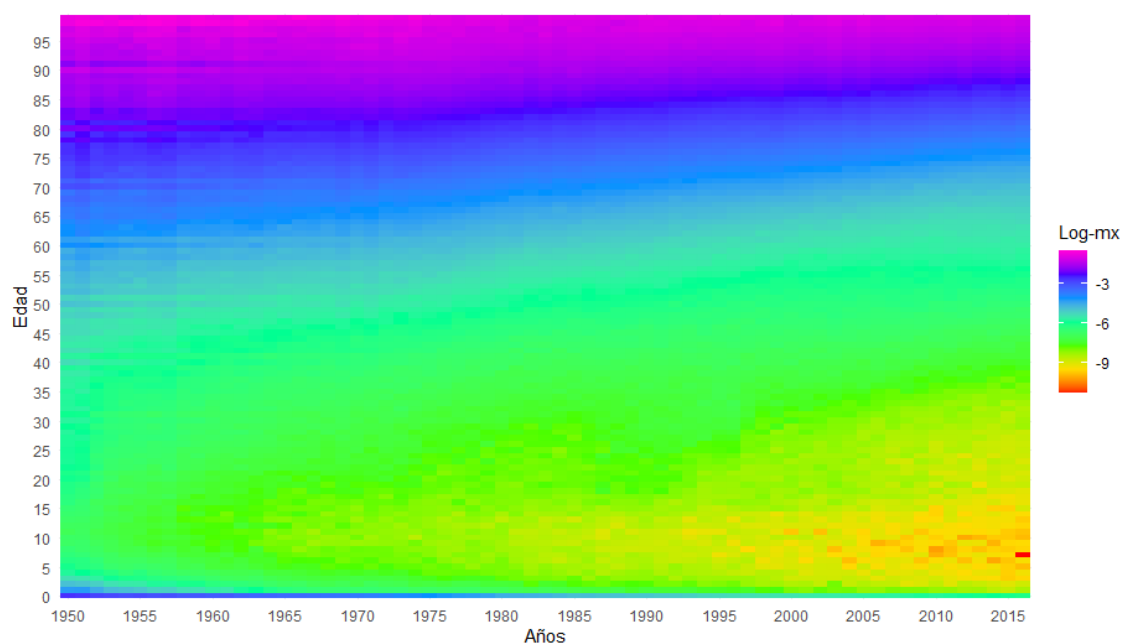


Figura 19. Mapa de calor de la log-mortalidad para el género femenino en España.

Los tonos fríos, como el azul y el violeta, representan una alta tasa de mortalidad (menor valor en términos logarítmicos), y tonos cálidos, como el amarillo y el naranja una baja tasa de mortalidad. Se observa como conforme avanzan los años, la mortalidad va concentrándose mayoritariamente en edades avanzadas, desde los 80 a los 90 años. Además, se aprecia que en una edad crítica como es la de un recién nacido, la mortalidad

también va disminuyendo con el paso de los años, aunque sigue siendo notablemente superior a edades próximas.

En cuanto al género masculino, se obtiene en la Figura 20 el mismo análisis. En este caso concreto, cabe destacar la disminución de la mortalidad con el paso del tiempo en las edades tempranas, aunque su tasa de supervivencia entre los 80 y 90 años sea algo inferior que en el caso del género femenino.

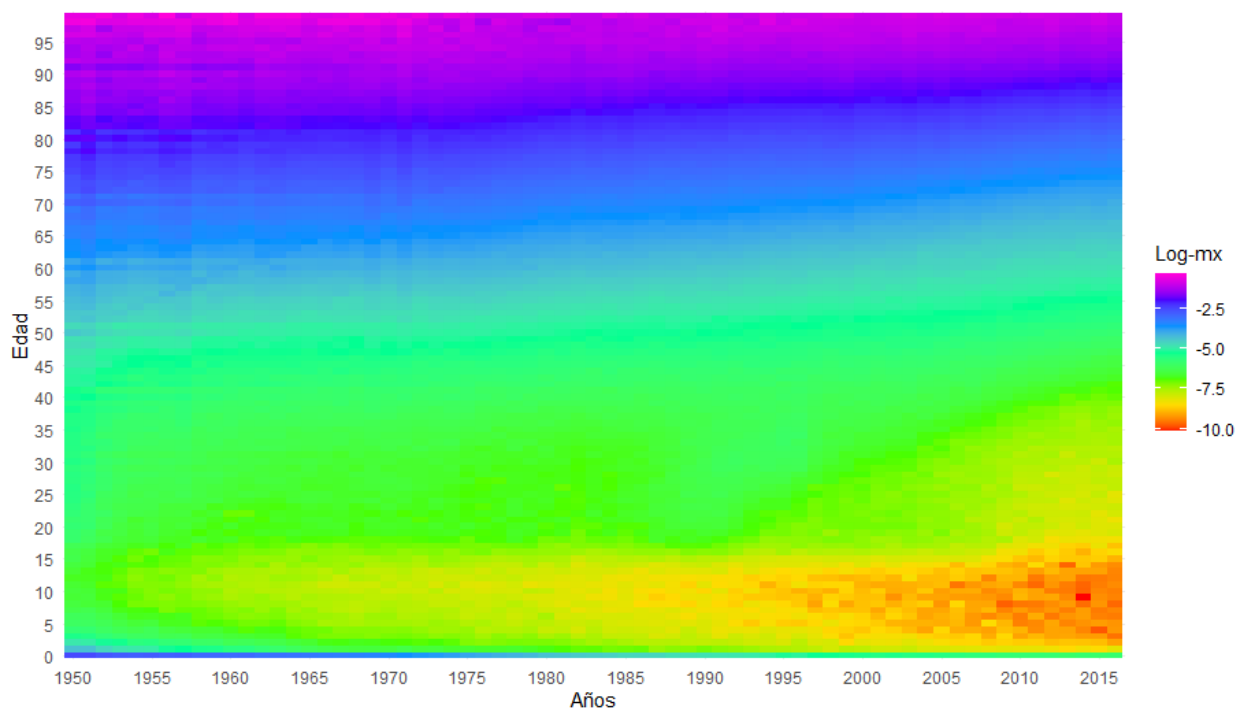


Figura 20. Mapa de calor de la log-mortalidad para el género masculino en España

### 6.1. Aplicación del Modelo de Lee-Carter

Siguiendo las pautas establecidas en el Capítulo 2, se ha procedido a calibrar los parámetros del modelo de Lee Carter mediante descomposición singular de valores. Los valores obtenidos a lo largo de las edades, para el caso de las variables  $\alpha$  y  $\beta$ , se pueden consultar con detalle en las figuras adjuntas a continuación.

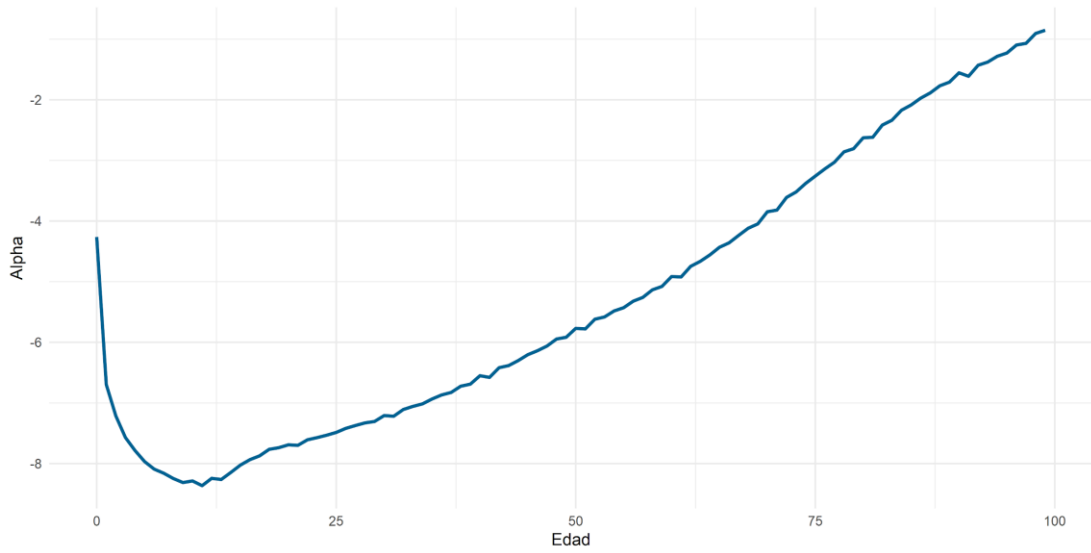


Figura 21. Representación del parámetro  $\alpha$  para el género femenino.

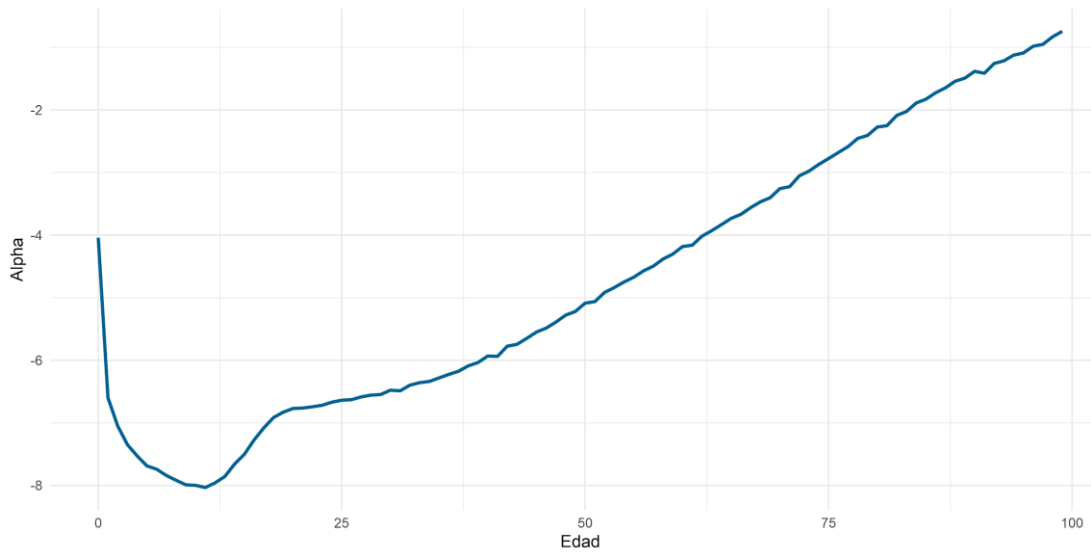


Figura 22. Representación del parámetro  $\alpha$  para el género masculino



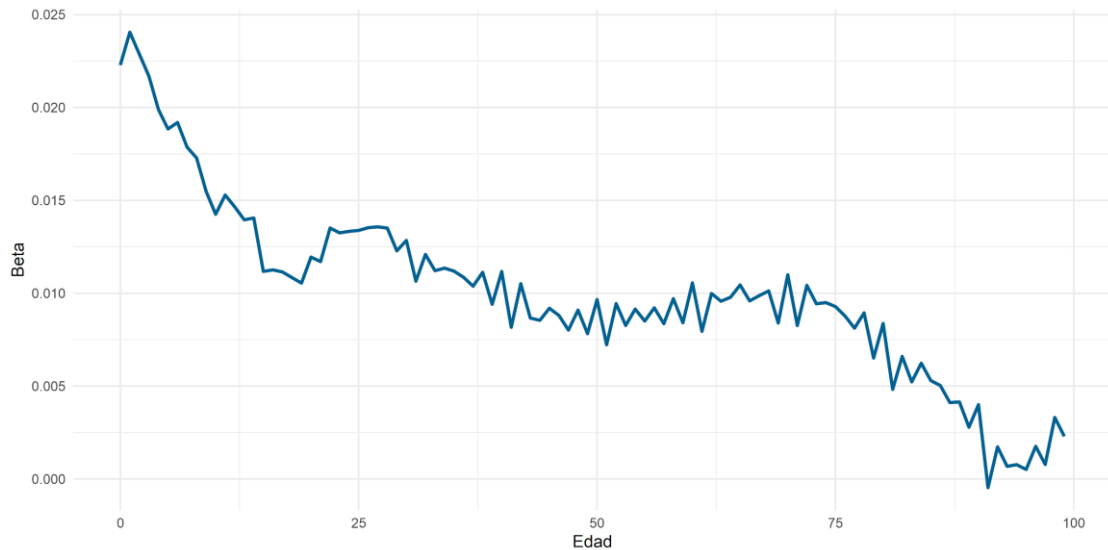


Figura 23. Representación del parámetro  $\beta$  para el género femenino.

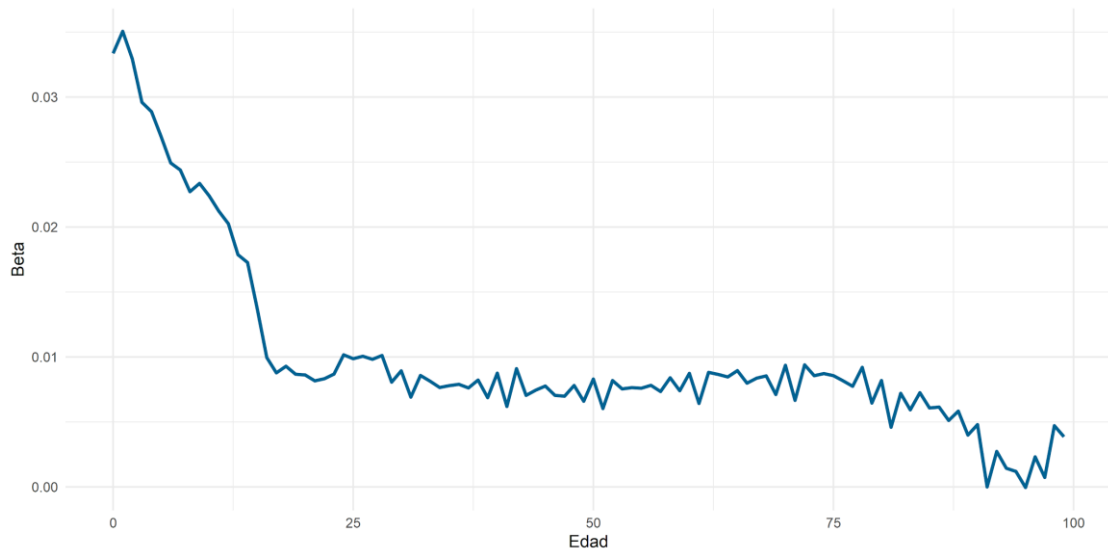


Figura 24. Representación del parámetro  $\beta$  para el género masculino.

Además, se estima el valor del componente temporal  $\kappa$ , y mediante el uso de la librería Forecast de R (Rob Hyndman et al., 2020), se proyecta como un paseo aleatorio con deriva hasta el último año de validación. Se puede observar el último año conocido, 2006, de color azul, mientras que los años de proyección (o validación) tienen un sombreado lila. En general, la tenencia temporal es decreciente y suavizada, y se observa que la proyección sigue dicho patrón para ambos géneros.

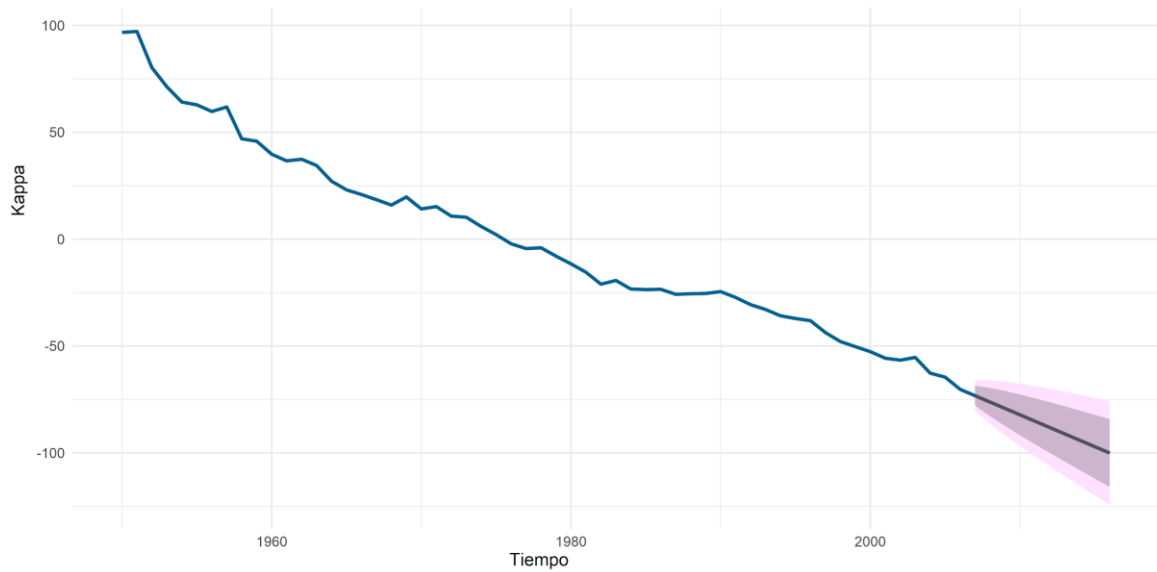


Figura 25. Representación del parámetro  $\kappa$  proyectado para el género femenino.

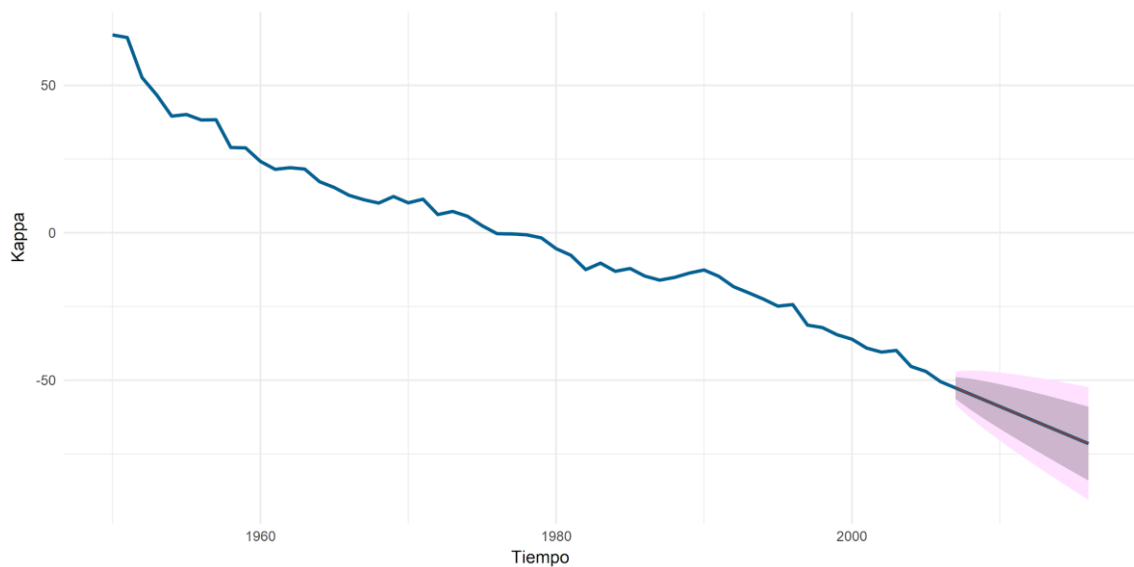


Figura 26. Representación del parámetro  $\kappa$  proyectado para el género masculino.

Finalmente, gracias a la proyección realizada, se obtienen los valores finales de la mortalidad y la log-mortalidad aplicando los parámetros de edad,  $\alpha$  y  $\beta$ , a la variable temporal  $\kappa$ .

Las predicciones finales a un año, cinco y diez de la mortalidad para el género femenino se muestran a continuación, tanto en su medida original como en términos logarítmicos, ya que, en edades tempranas, al ser los valores centésimas de unidad no se aprecia gráficamente la diferencia.

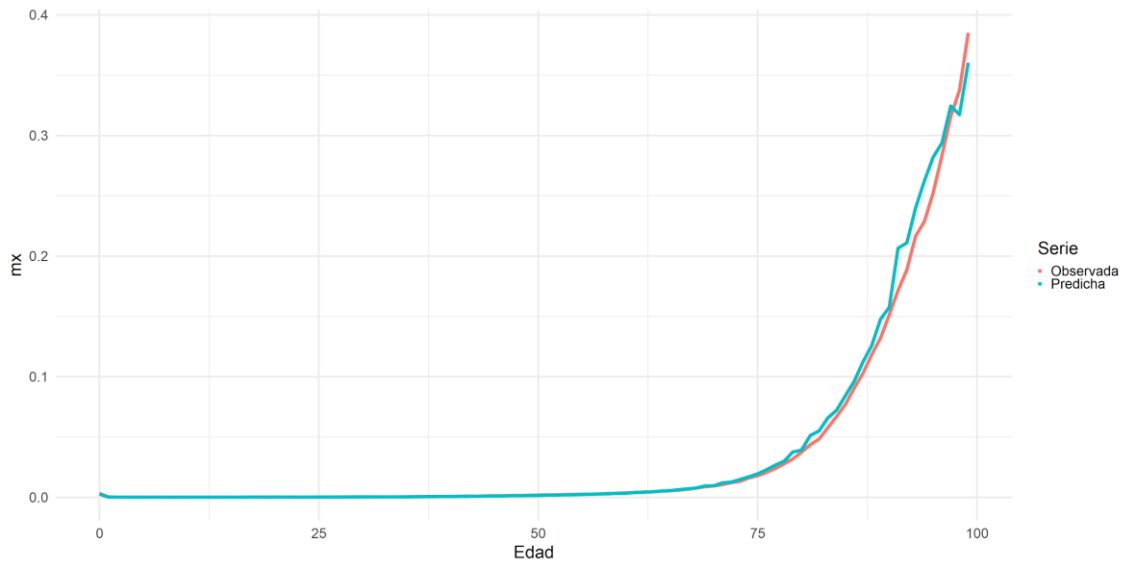


Figura 27. Mortalidad observada y predicha en el año 2007 para el género femenino mediante LC.

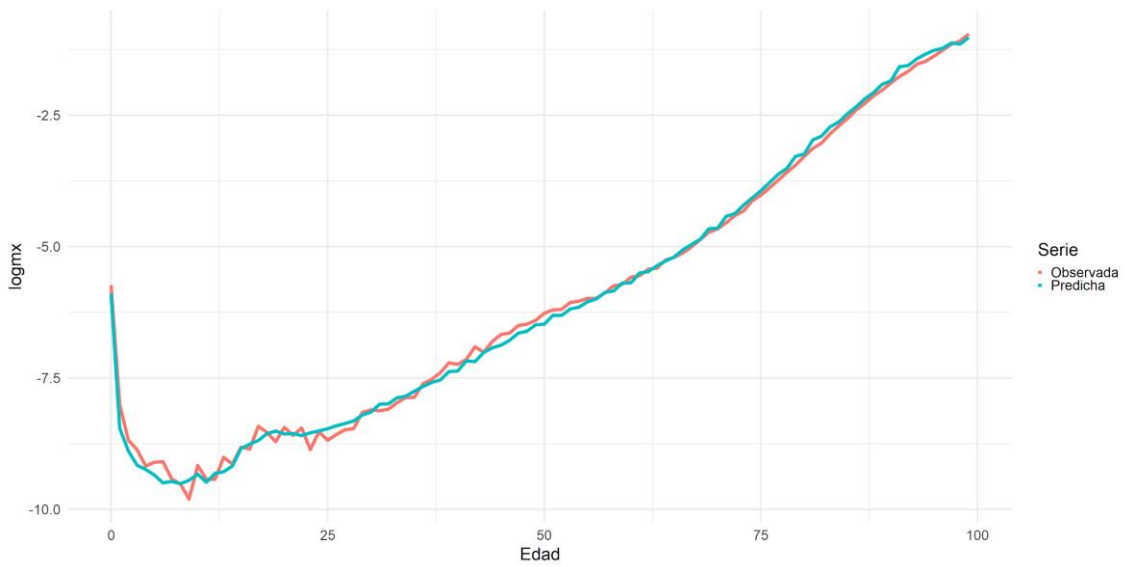


Figura 28. Log-Mortalidad observada y predicha en el año 2007 para el género femenino mediante LC.

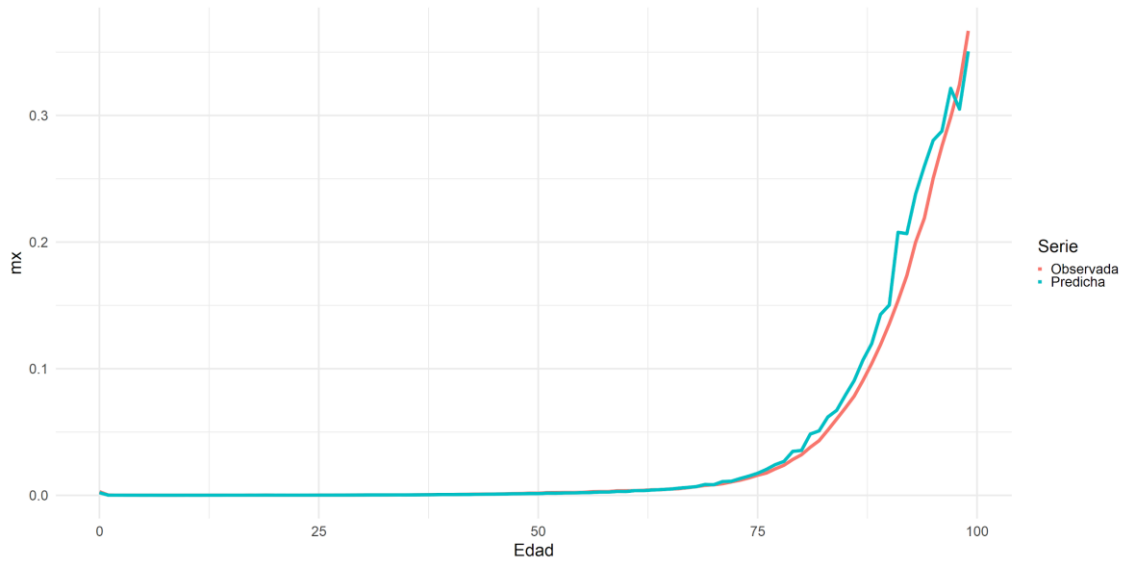


Figura 29. Mortalidad observada y predicha en el año 2011 para el género femenino mediante LC.

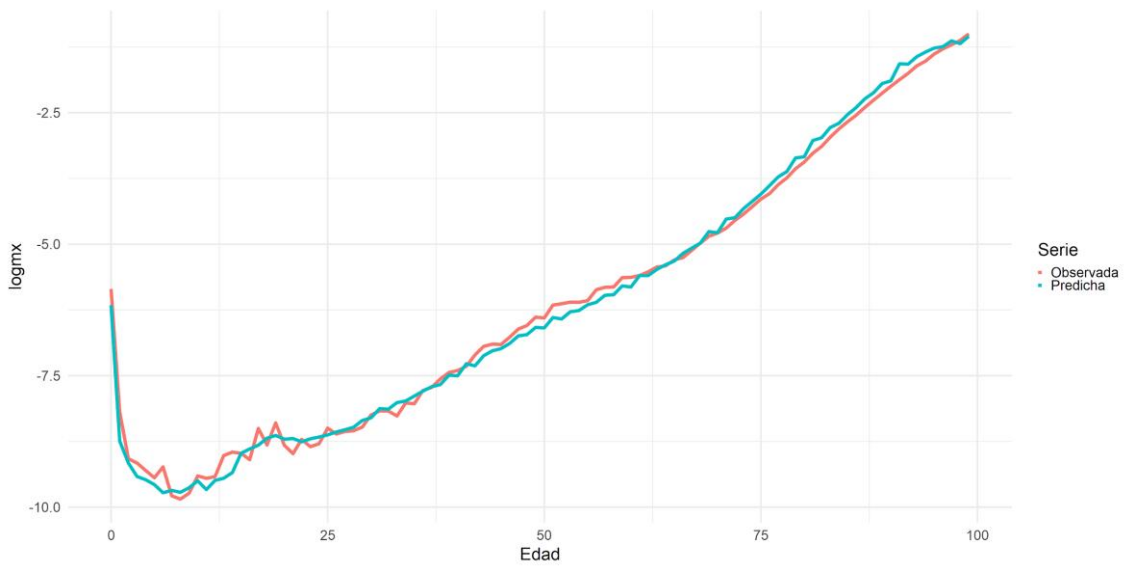


Figura 30. Log-Mortalidad observada y predicha en el año 2011 para el género femenino mediante LC.

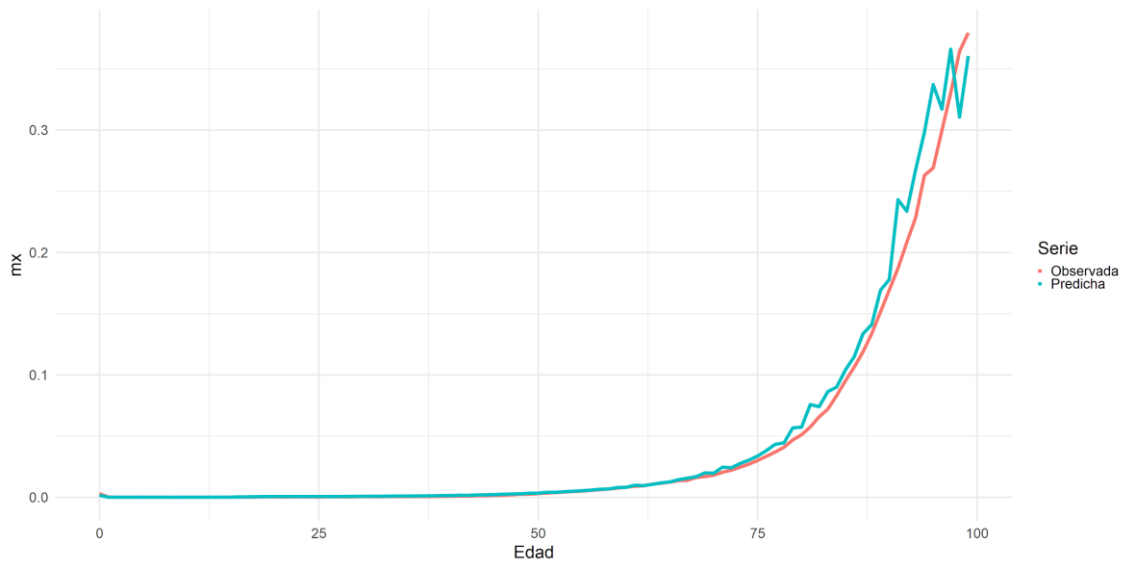


Figura 31. Mortalidad observada y predicha en el año 2016 para el género femenino mediante LC.

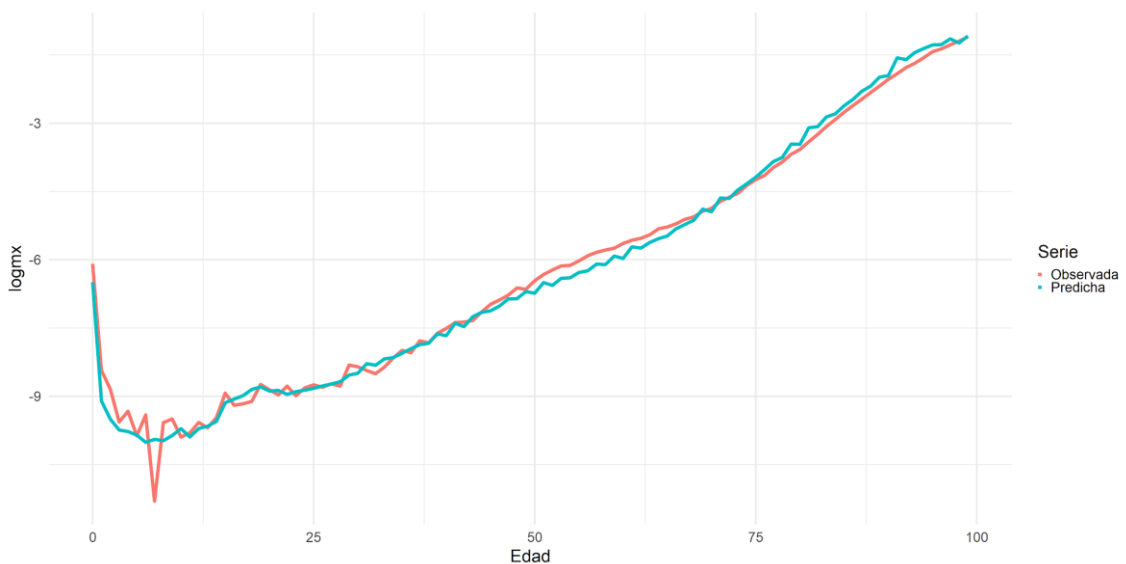


Figura 32. Log-Mortalidad observada y predicha en el año 2016 para el género femenino mediante LC.

Gráficamente, se observa que la predicción es más acertada en periodos temporales cortos (1 y 5 años), ya que la predicción a más largo plazo (10 años) se desvía notablemente en el tramo de edad de 65 a 100 años, aunque bien es cierto que en los tres casos el modelo no es precisamente acertado en las edades mencionadas.

Para poder comprobar con certeza esta observación, se ha extraído para los tres años una medida estándar de evaluación de predicciones como es el error cuadrático medio (MSE), y el error porcentual medio absoluto (MAPE), utilizado frecuentemente en este tipo de evaluación de pronósticos debido a su fácil interpretación.

Tabla 2. Evaluación de la predicción mediante LC para el género femenino.

	<b>MSE</b>	<b>MAPE</b>
<b>LC 2007 <math>m_{x,t}</math></b>	6.282814e-05	0.1139697
<b>LC 2011 <math>m_{x,t}</math></b>	0.000113211	0.1466
<b>LC 2016 <math>m_{x,t}</math></b>	0.0001518404	0.185223

Tal y como se preveía, en la Tabla 2 se confirma que conforme más extenso es el horizonte temporal, peor es la predicción. En términos porcentuales, mientras que, en el año 2007, la predicción para el género femenino mediante el modelo de Lee Carter fallaba por término medio un 11,39%, en 2016 fallaba un 18,52%, lo cual es notablemente superior y muestra una de las deficiencias señaladas por diferentes autores en este modelo.

Se ha realizado el mismo análisis para el género masculino en aras de poder comprobar si las conclusiones obtenidas anteriormente se repiten en este caso. A continuación, se muestran las predicciones de la mortalidad en su medida original y en términos logarítmicos.

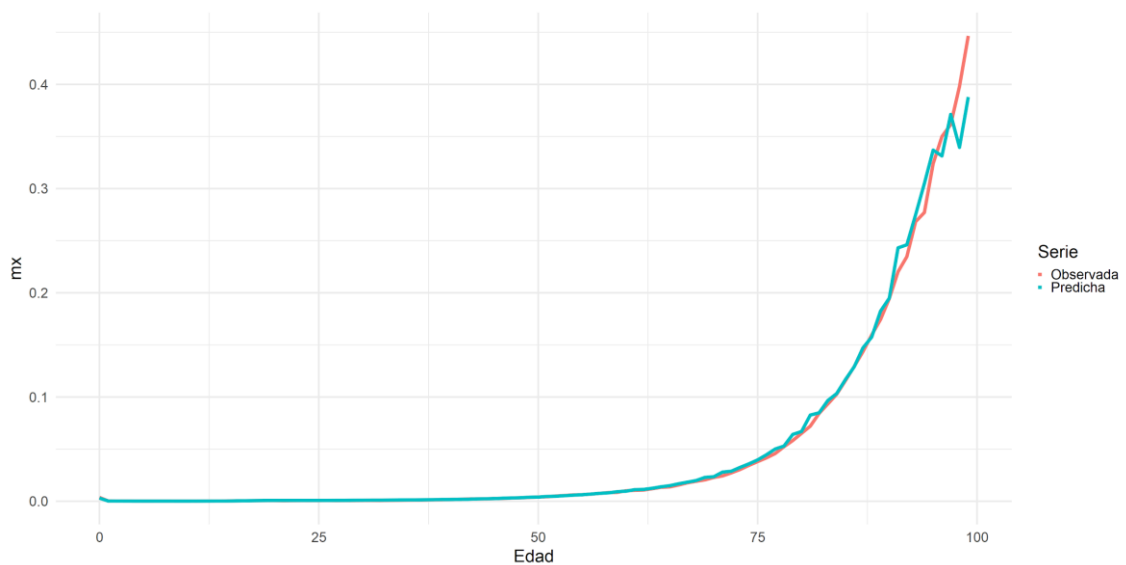


Figura 33. Mortalidad observada y predicha en el año 2007 para el género masculino mediante LC.

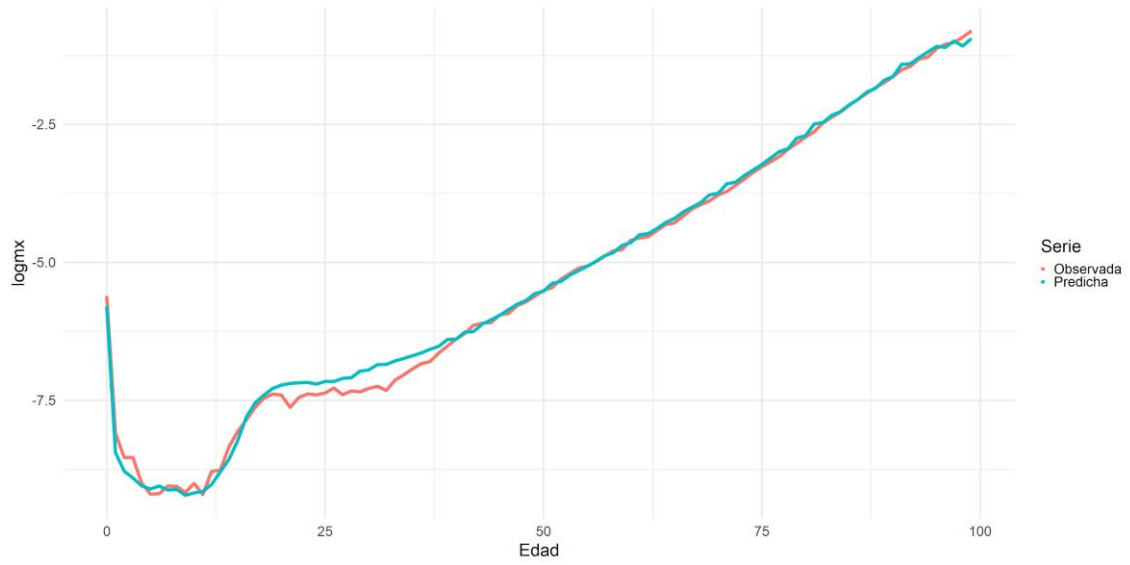


Figura 34. Log-Mortalidad observada y predicha en el año 2007 para el género masculino mediante LC.

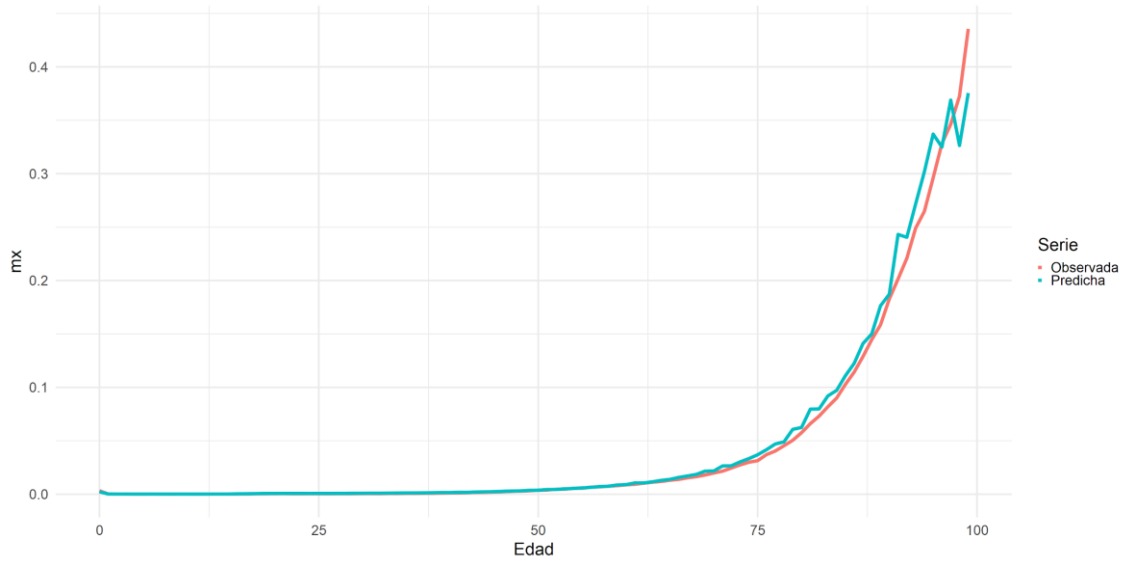


Figura 35. Mortalidad observada y predicha en el año 2011 para el género masculino mediante LC.

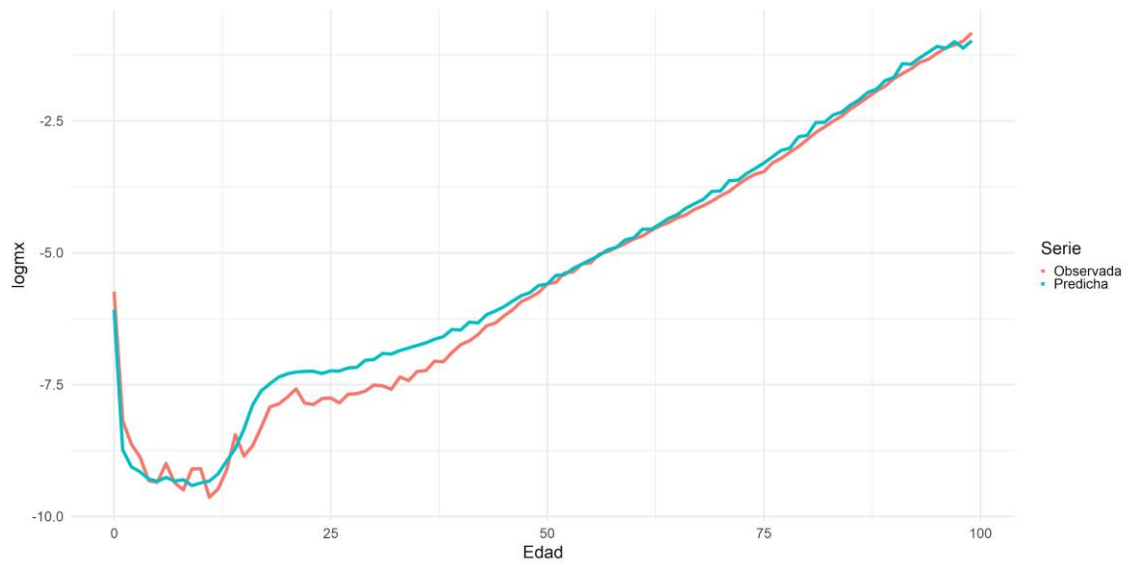


Figura 36. Log-Mortalidad observada y predicha en el año 2011 para el género masculino mediante LC.

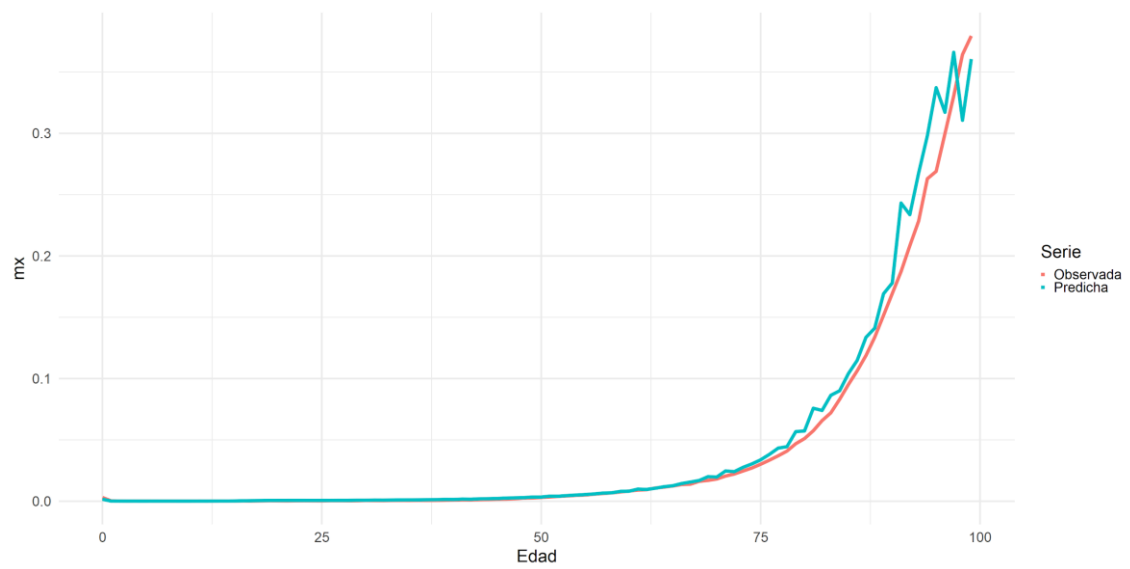


Figura 37. Mortalidad observada y predicha en el año 2016 para el género masculino mediante LC.



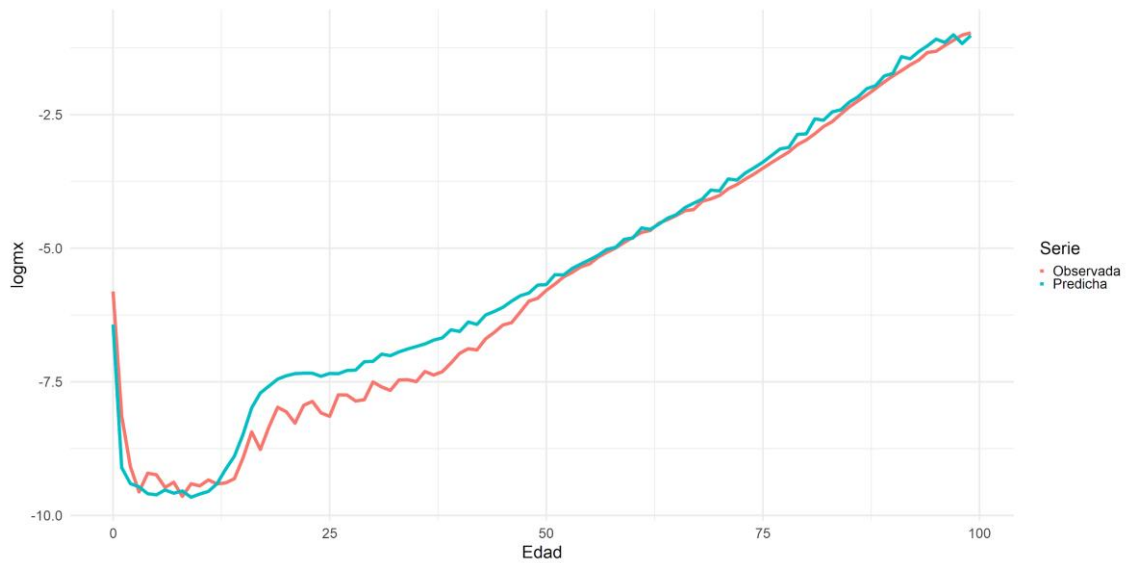


Figura 38. Log-Mortalidad observada y predicha en el año 2016 para el género masculino mediante LC.

Como ocurría con en el caso anterior, la calidad de la predicción decrece conforme se aumenta el horizonte temporal de predicción. Además, cabe destacar que, en el caso del género masculino, la predicción es aún menos acertada a 5 y 10 años, con un error medio porcentual absoluto de 20,55% y 24,52% respectivamente (observable en la Tabla 3). Esto puede deberse principalmente a que la población masculina ha sufrido mayores fluctuaciones en su demografía a lo largo del tiempo, y el modelo de Lee Carter realiza una predicción basada en la calibración de sus parámetros mediante datos históricos. En consecuencia, la predicción posterior es menos exacta para este género.

Tabla 3. Evaluación de la predicción mediante LC para el género masculino.

	<b>MSE</b>	<b>MAPE</b>
<b>LC 2007 <math>m_{x,t}</math></b>	9.250154e-05	0.1050282
<b>LC 2011 <math>m_{x,t}</math></b>	0.0001331223	0.205548
<b>LC 2016 <math>m_{x,t}</math></b>	0.0001775664	0.2462542

## 6.2. Aplicación de la red neuronal recurrente de tipo LSTM

Una vez analizado el modelo de Lee Carter y sus resultados, los cuales servirán como

medida de comparación, se procede a realizar el objetivo principal de este trabajo, el cuál es la predicción de la mortalidad tanto del género femenino como del masculino mediante una Red Neuronal Recurrente de tipo LSTM. Para ello, se han utilizado las librerías Keras (Falbel, Allaire, Chollet et al., 2020) y Tensorflow de R (Falbel, Allaire, Yuan Tang et al., 2020)

### **6.2.1. Elección de los hiper parámetros.**

Anteriormente, se han explicado con detalle conceptos básicos como el de neurona, capa o el funcionamiento matemático de la optimización de pesos y se ha profundizado más aún en la Red Neuronal Recurrente de tipo LSTM, su estructura interna y su modo de proceder. En este apartado, se ponen en práctica dichos conceptos, teniendo que realizar una serie de decisiones esenciales para el desarrollo a posteriori de la red neuronal.

Los hiper parámetros son las variables que determinan la estructura de la red, como el número de capas o de neuronas ocultas, y las variables que determinan cómo se entrena la red, como la tasa de aprendizaje. Para la elección de los hiper parámetros, no existe una regla fija o un procedimiento generalizado a seguir. Normalmente, lo indicado es la elección de un valor inicial, y a partir de esa configuración ir probando alternativas que puedan ser acertadas y que mejoren el desempeño de la red neuronal de forma general.

#### *- Estructura de capas:*

El paquete Keras utilizado funciona de forma que se necesitan ensamblar capas para construir modelos. Por ello, en primer lugar, se han ensamblado tres capas descritas a continuación:

- Una capa de entrada o input, la cual recibirá los datos iniciales para su procesamiento por las capas subsiguientes. Esta capa es el comienzo del flujo de trabajo de la red descrita.
- Una capa oculta LSTM, que recoge toda la metodología explicada con anterioridad. Las funciones de activación utilizadas son las mismas que describen los autores en su modelo original.
- Una capa densa de salida. Simplemente es una capa clásica de redes neuronales, pero totalmente conectada, donde cada nodo de entrada está conectado a cada nodo de salida de la capa anterior.

- *Estandarización:*

Dado que hay valores extremos en los datos que pueden alterar la calibración de los pesos, como puede ser la tasa de mortalidad a edades avanzadas, se hace precisa la utilización de una técnica de preprocesamiento de datos como es el MinMaxScaler. Este simplemente escala o estandariza cada valor de una matriz de datos a un rango dado. Su expresión se da de la siguiente forma:

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Donde  $\min(x)$  y  $\max(x)$  son los valores mínimo y máximo de la variable  $x$  dado su rango.

- *Función de pérdidas y optimizador:*

El propósito de las funciones de pérdida es calcular la cantidad de error que un modelo debe tratar de minimizar durante la fase de entrenamiento. En este caso se elige como función de pérdidas el error cuadrático medio. Como se comentó en capítulos anteriores, éste calcula la diferencia entre lo observado y lo predicho.

Como optimizador se ha implementado el algoritmo de Adam como una alternativa al descenso del gradiente básico. Este es un método de descenso de gradiente, pero estocástico, que se basa en la estimación adaptativa de momentos de primer y segundo orden. Está diseñado específicamente para el entrenamiento de redes neuronales profundas. Según los autores el método es "eficiente desde el punto de vista computacional, tiene pocos requisitos de memoria, es invariable a la reescalada diagonal de los gradientes, y es muy adecuado para problemas que son grandes en términos de datos/parámetros" (Kingma & Ba, 2014).

- *Preparación de los datos de entrada para redes LSTM:*

A la hora de introducir los datos en la capa de entrada de la red neuronal, es necesario entender que utilizar el formato original de los datos no es factible. Las redes neuronales LSTM precisan de una entrada en formato matricial 3D, donde las dimensiones son las siguientes:

- Batches o lotes, que serían muestras de la información.

- Time-steps, que sería un punto de observación hacia atrás en la muestra.
- Observaciones por cada time-step.

Por ejemplo, en el caso de que se eligiese time-steps de 10, que sería semejante a observar hacia atrás 10 periodos temporales, considerando el conjunto de entrenamiento desde 1950 hasta 2006 (57 años), las observaciones totales serían de  $(57-10) * 100 = 4700$ . La matriz quedaría definida en sus dos primeras dimensiones (2D) de la siguiente forma:

$x_{1,1950}$	$x_{1,1951}$	...	$x_{1,1958}$	$x_{1,1959}$
$x_{2,1950}$	$x_{2,1951}$	...	$x_{2,1958}$	$x_{2,1959}$
...	...	...	...	...
$x_{100,1950}$	$x_{100,1951}$	...	$x_{100,1958}$	$x_{100,1959}$
$x_{1,1951}$	$x_{1,1952}$	...	$x_{1,1959}$	$x_{100,1960}$
...	...	...	...	...
$x_{1,1996}$	$x_{1,1997}$	...	$x_{1,2004}$	$x_{1,2005}$
...	...	...	...	...
$x_{100,1996}$	$x_{100,1997}$	...	$x_{100,2004}$	$x_{100,2005}$

obteniendo una dimensión de 4700 filas y 10 columnas.

Añadiendo una tercera dimensión que sería el número de muestras, si elegimos la cantidad de 5, por ejemplo, para un individuo de 65 años estaría entrando a la capa de entrada la siguiente muestra las edades:

$x_{63}$	$x_{64}$	$x_{65}$	$x_{66}$	$x_{67}$
----------	----------	----------	----------	----------

en sus respectivos periodos temporales.

Para el caso de la primera o última muestra, cómo serían las que contemplan a individuos de edades 1 o 100, se repiten estos vectores a la izquierda o a la derecha, hasta por cumplir el lote. Por ejemplo, en el caso de un individuo de edad 1:

$x_1$	$x_1$	$x_1$	$x_2$	$x_3$
-------	-------	-------	-------	-------

Y para un individuo de edad 100:

$x_{98}$	$x_{99}$	$x_{100}$	$x_{100}$	$x_{100}$
----------	----------	-----------	-----------	-----------

- *Neuronas:*

En cuanto al número de neuronas elegidas en la capa oculta LSTM, se han hecho sucesivas pruebas para llegar a un resultado consistente. Estas han consistido en partir de una configuración inicial e ir probando diferentes variaciones que disminuyan el error de la predicción.

En las Tablas 4 y 5 se muestran las configuraciones más reseñables que se han podido obtener para ambos géneros, y sombreado de color gris se destaca aquella que minimiza de forma más óptima el error cuadrático medio de la predicción. También se muestra el error medio porcentual absoluto a título informativo, ya que mediante esta tabla podemos observar la calidad predictiva de la red en su conjunto y de forma individualizada para la configuración escogida. Como configuración se entiende elegir un número adecuado de time steps (Ts), lotes ( $x_t$ ) y neuronas de la capa oculta LSTM (N).

Cabe reseñar que no siempre una misma prueba repetida numerosas veces va a reportar el mismo resultado, ya que el umbral de la red es inicializado de forma aleatoria. Por ello, la tabla mostrada es la media de 100 pruebas realizadas para cada configuración.

Tabla 4. Pruebas de hiper parámetros para el género femenino.

Hiper Parámetros	2007		2011		2016	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
Ts = 5 $x_t$ = 3 N = 10	7.345668e-05	0.08021887	4.591032e-05	0.08684031	8.702309e-05	0.08661209
Ts = 10 $x_t$ = 3	1.930411e-05	0.07577357	7.855142e-06	0.08534915	1.717763e-06	0.08427038

<b>N = 15</b>						
<b>Ts = 15</b>						
$x_t = 3$	4.145834e-05	0.08034152	2.458096e-05	0.07839657	5.707554e-06	0.08124261
<b>N = 15</b>						
<b>Ts = 5</b>						
$x_t = 5$	7.699673e-06	0.0723775	5.342539e-06	0.08395546	4.767781e-06	0.08457121
<b>N = 10</b>						
<b>Ts = 10</b>						
$x_t = 5$	6.877722e-06	0.06534018	5.243048e-06	0.0758652	9.704619e-06	0.08478406
<b>N = 20</b>						
<b>Ts = 10</b>						
$x_t = 7$	7.895413e-05	0.08582038	7.3607e-05	0.09004085	9.280664e-05	0.101203
<b>N = 25</b>						
<b>Ts = 10</b>						
$x_t = 5$	1.496376e-05	0.07835985	9.25668e-06	0.0910139	1.686914e-05	0.08832415
<b>N = 30</b>						
<b>Ts = 20</b>						
$x_t = 7$	1.692733e-05	0.08651771	1.443564e-05	0.09732994	2.87812e-05	0.09473947
<b>N = 20</b>						
<b>Ts = 10</b>						
$x_t = 7$	2.080158e-05	0.07383038	1.054174e-05	0.0881346	2.743455e-06	0.09068737
<b>N = 20</b>						
<b>Ts = 5</b>						
$x_t = 7$	8.784474e-06	0.0866182	9.120614e-06	0.0866182	2.687971e-06	0.08106353
<b>N = 20</b>						

Tabla 5. Pruebas de hiper parámetros para el género masculino.

	<b>2007</b>		<b>2011</b>		<b>2016</b>	
<b>Hiper Parámetros</b>	<b>MSE</b>	<b>MAPE</b>	<b>MSE</b>	<b>MAPE</b>	<b>MSE</b>	<b>MAPE</b>
<b>Ts = 5</b>						
$x_t = 3$	2.989978e-05	0.07126891	4.156272e-05	0.09733853	8.353549e-05	0.11311
<b>N = 10</b>						
<b>Ts = 10</b>						
$x_t = 3$	1.922754e-05	0.06884481	2.142741e-05	0.08823529	1.660572e-05	0.1041833
<b>N = 15</b>						
<b>Ts = 15</b>						
$x_t = 3$	3.666165e-05	0.05127295	2.291181e-05	0.06814234	1.596965e-05	0.08053234
<b>N = 15</b>						
<b>Ts = 5</b>						
$x_t = 5$	1.132598e-05	0.0600082	1.050972e-05	0.07340844	8.237059e-06	0.09608338
<b>N = 10</b>						
<b>Ts = 10</b>						
$x_t = 5$	1.331669e-05	0.05459359	1.230451e-05	0.06824254	1.530955e-05	0.07219415
<b>N = 20</b>						

<b>Ts = 10</b>						
$x_t = 7$	1.423267e-05	0.048696	2.098736e-05	0.07522989	2.00072e-05	0.07346611
<b>N = 25</b>						
<b>Ts = 10</b>						
$x_t = 5$	2.110143e-05	0.05429299	1.24748e-05	0.06122804	6.042789e-06	0.06831325
<b>N = 30</b>						
<b>Ts = 20</b>						
$x_t = 7$	1.164643e-05	0.06070023	1.429723e-05	0.08219199	2.268197e-05	0.09810049
<b>N = 20</b>						
<b>Ts = 10</b>						
$x_t = 7$	2.324396e-05	0.05474143	4.622376e-05	0.07900502	6.114299e-05	0.09138427
<b>N = 20</b>						
<b>Ts = 5</b>						
$x_t = 7$	1.57087e-05	0.06099751	2.165277e-05	0.08375353	6.225494e-05	0.1011127
<b>N = 20</b>						

Se concluye que para ambos géneros la configuración  $\{Ts = 5, x_t = 5, N = 10\}$  es la más adecuada debido a que el error cuadrático medio es notablemente inferior en predicciones a 10 años, mientras que a un menor horizonte temporal es similar al resto de configuraciones.

- *Entrenamiento:*

En el entrenamiento de las redes neuronales, la elección del número de epochs (como ya se explicó con anterioridad, un ciclo compuesto de ir hacia delante para calcular la predicción y su error con respecto a lo observado, y luego hacia atrás para corregirlo y optimizar) es una cuestión a tratar con delicadeza.

La utilización de demasiados epochs puede llevar a un sobreajuste del modelo o red neuronal, mientras que utilizar pocos ciclos puede llevar a un subajuste. En Machine Learning, el uso de la regla de parada temprana o “early stopping rule” es una forma de evitar el sobreajuste o subajuste cuando se entrena la red mediante un método iterativo (aquí podría englobarse el método del descenso del gradiente y todas sus variaciones más complejas). La regla de parada temprana proporciona una guía de cuántas iteraciones se pueden realizar antes de que la red comience a sobre ajustarse. Su funcionamiento es simple, se detendrá el entrenamiento una vez que el rendimiento del modelo deja de mejorar comparándolo con el conjunto de datos de validación. En la Figura 39 se puede observar de forma sencilla una representación gráfica de este método.

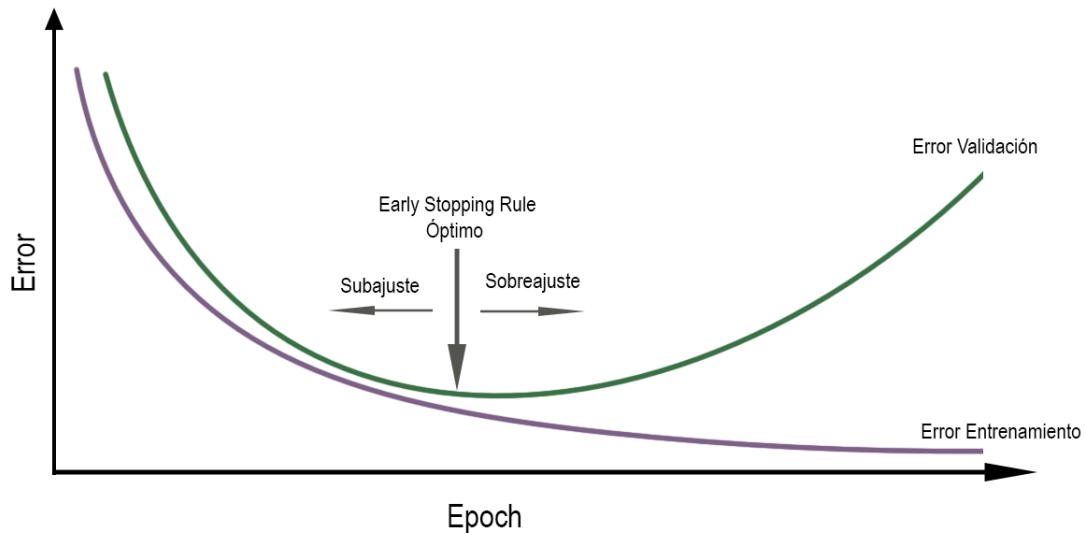


Figura 39. Representación del sobreajuste y subajuste.

Se puede identificar un buen ajuste en la red cuando el error de entrenamiento y validación disminuyen consistentemente hasta un punto de estabilidad en el que no existe una gran brecha entre los dos valores (Brownlee, 2019). De forma general, el error será menor en el conjunto de datos de entrenamiento que en el de validación. Por ello, es normal encontrar alguna leve diferencia entre ambas curvas.

En la Figura 40 se muestra una representación del error obtenido conforme aumenta el número de epochs para la configuración de red finalmente elegida. En general, no parece que haya un problema de sobreajuste o subajuste y el modelo ha convergido correctamente. Esto podemos comprobarlo ya que los errores disminuyen hasta un punto de estabilidad, donde existe una pequeña brecha entre los errores del conjunto de entrenamiento y de validación, siendo el primero mencionado ligeramente inferior.



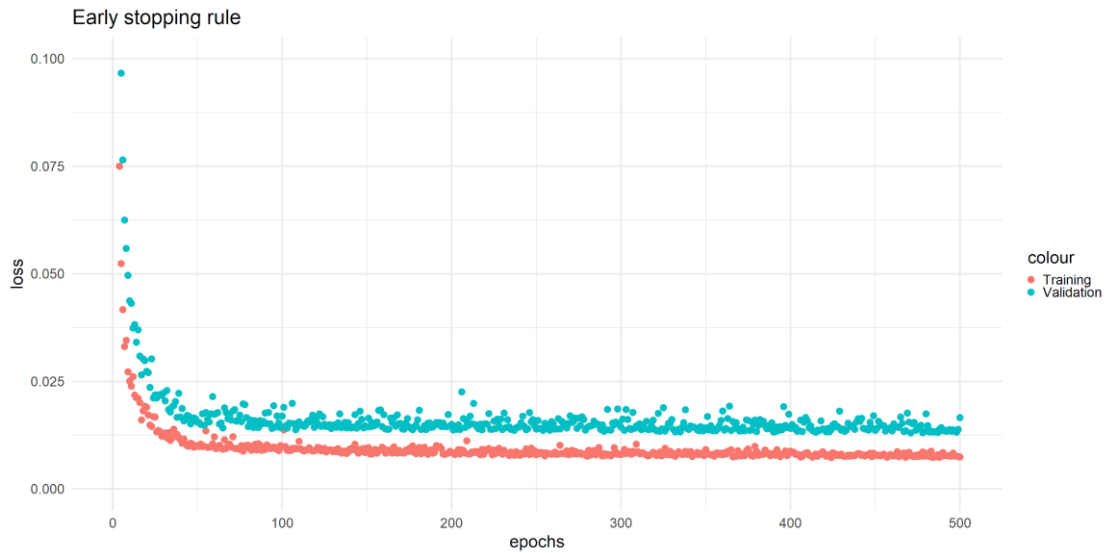


Figura 40. Representación de la regla de parada temprana en el modelo elegido.

### 6.2.2. Resultados

A continuación, se exponen los resultados obtenidos mediante la red neuronal recurrente de tipo LSTM para la predicción de la mortalidad en el género femenino.

En las Figuras 41 y 42, se observa la alta precisión a la hora de generar las series de mortalidad, obteniendo para edades tempranas unos valores prácticamente similares a los originales. La casuística que se encontraba en el modelo de Lee Carter para edades avanzadas, en las que la predicción empezaba a fallar a partir de los 65 años, no se observa mediante el uso de la red neuronal, aunque bien es cierto que el error empieza a aumentar ligeramente a partir de los 90 años. En general, se podría decir que, para predicciones a un horizonte temporal de un año, este tipo de red tiene un rendimiento sobresaliente en todos los tramos de edades.

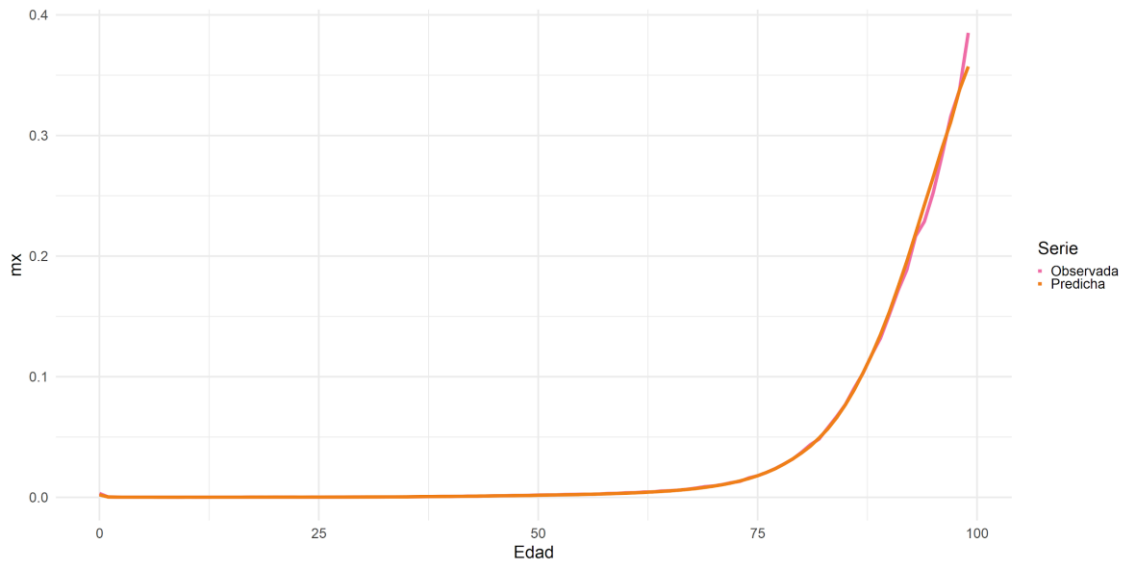


Figura 41. Mortalidad observada y predicha en el año 2007 para el género femenino mediante LSTM.

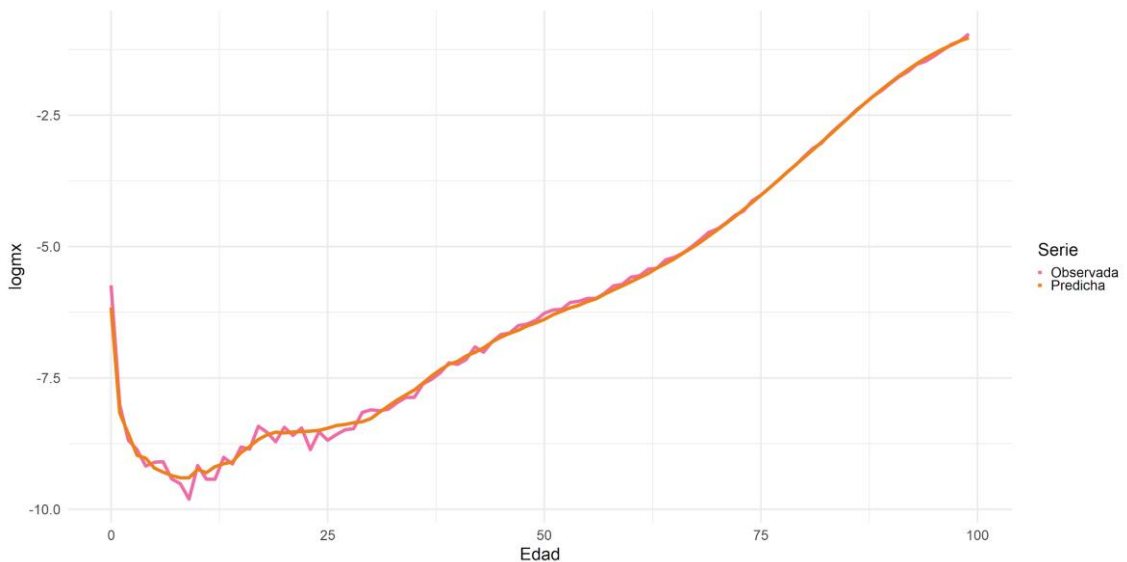


Figura 42. Log-Mortalidad observada y predicha en el año 2007 para el género femenino mediante LSTM.

Posteriormente, se procede a realizar las mismas predicciones a 5 y 10 años en aras de realizar una comparación consistente con el modelo de Lee Carter. En las Figuras 43 y 44, se aprecia que la red neuronal artificial sigue realizando una predicción muy cercana a los datos reales para el año 2011, repitiendo el mismo patrón anteriormente visto de un mayor acierto en edades tempranas, y una leve desviación en edades muy avanzadas (posteriores a los 90 años).

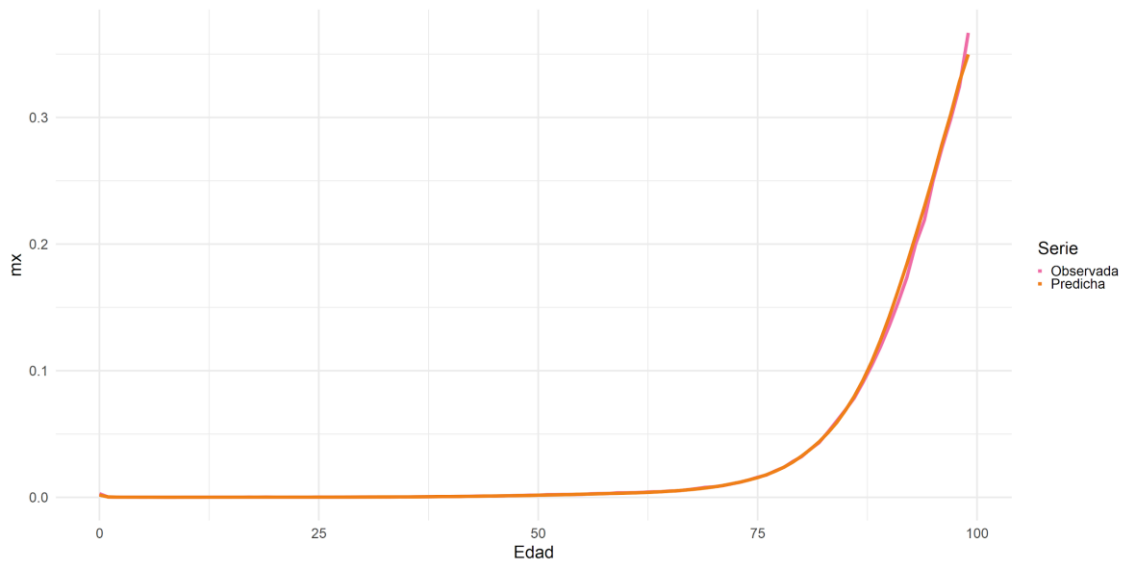


Figura 43. Mortalidad observada y predicha en el año 2011 para el género femenino mediante LSTM.

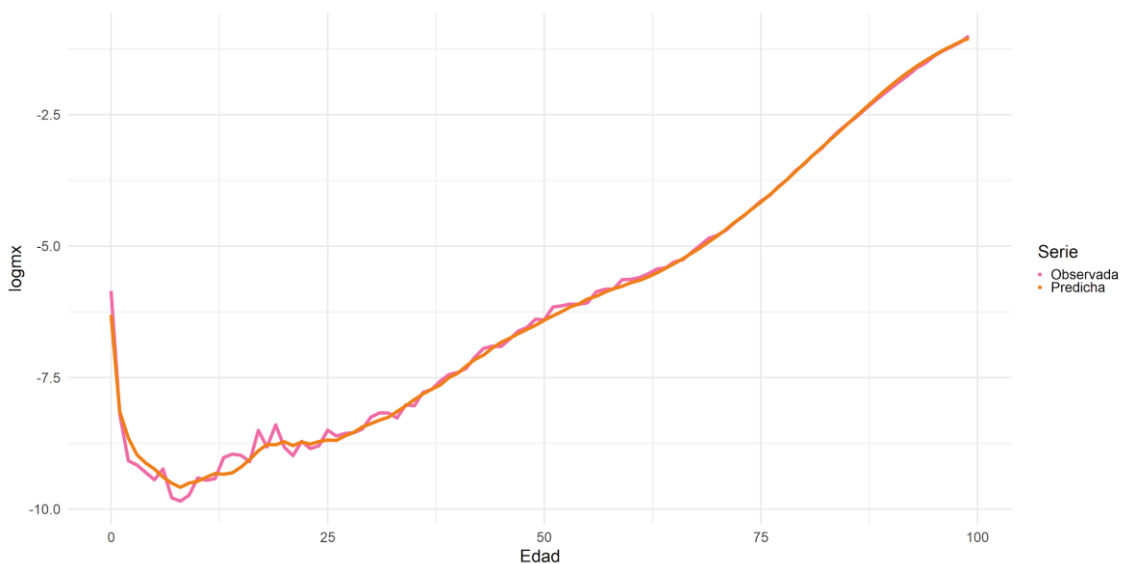


Figura 44. Log-Mortalidad observada y predicha en el año 2011 para el género femenino mediante LSTM.

Sorprendentemente, en las Figuras 45 y 46, se observa que la red neuronal sigue prediciendo con una gran ratio de acierto incluso en proyecciones a 10 años, difiriendo notablemente de lo obtenido con el modelo paramétrico de Lee Carter para el género femenino. Cabe destacar que este escenario era de esperar en cierta medida, ya que esta tipología concreta de red neuronal artificial tiene memoria tanto a largo como a corto plazo, permitiendo que ella misma descubra patrones de comportamiento, eliminando toda aquella información irrelevante y guardando toda aquella que considere imprescindible para su posterior uso, tareas que un modelo paramétrico no puede llegar a realizar para mejorar su precisión.

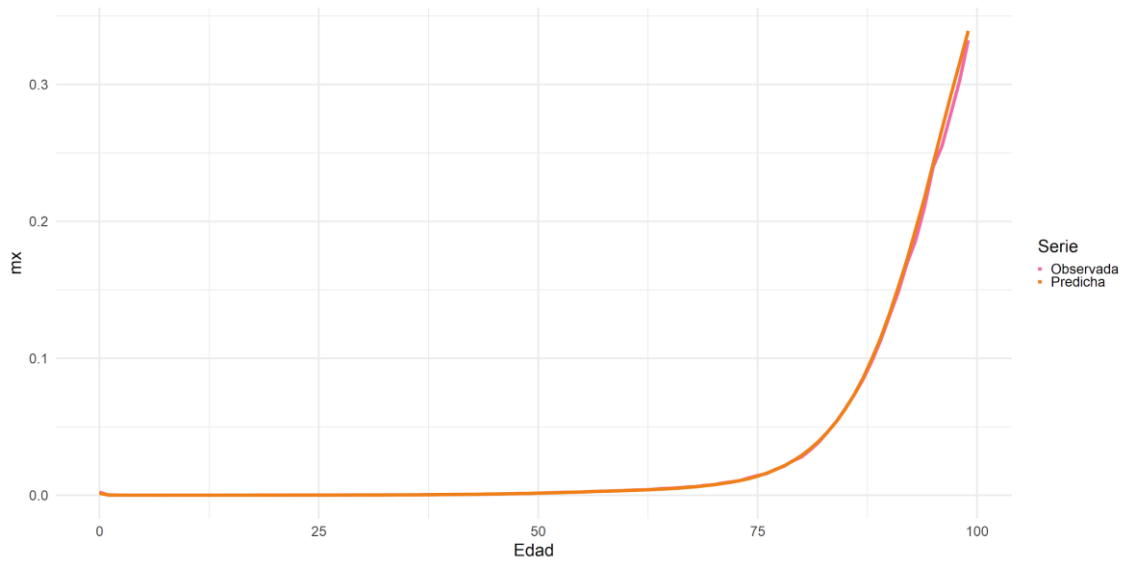


Figura 45. Mortalidad observada y predicha en el año 2016 para el género femenino mediante LSTM.

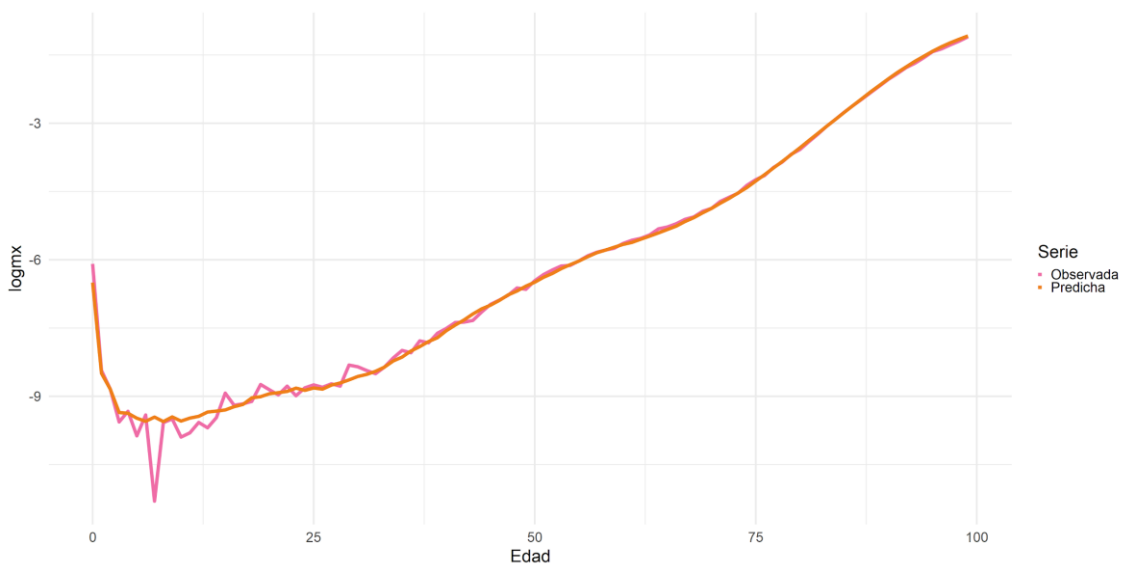


Figura 46. Log-Mortalidad observada y predicha en el año 2016 para el género femenino mediante LSTM.

Finalmente, en la Tabla 6 podemos observar en detalle el análisis del error cuadrático medio y el error porcentual medio absoluto para el género femenino. Efectivamente, la calidad de la predicción no mejora conforme se va extendiendo el horizonte de proyección, rompiendo todos los esquemas de que a mayor horizonte temporal peor es la predicción, patrón observado en multitud de modelos de mortalidad paramétricos conocidos.

Concretamente, hablamos de un error cuadrático medio de mil millonésimas para las tres series que no varía notablemente conforme pasa el tiempo. En términos

porcentuales, se hablaría de un error porcentual medio de aproximadamente un 8%, que estará concentrado habitualmente en los tramos más superiores de edades.

Tabla 6. Evaluación de la predicción mediante LSTM para el género femenino.

	MSE	MAPE
<b>LSTM 2007 <math>m_{x,t}</math></b>	7.699673e-06	0.0723775
<b>LSTM 2011 <math>m_{x,t}</math></b>	5.342539e-06	0.08395546
<b>LSTM 2016 <math>m_{x,t}</math></b>	4.767781e-06	0.08457121

En las figuras mostradas seguidamente, se realiza el mismo análisis para el género masculino, y se representan las proyecciones para los años 2007, 2011 y 2016.

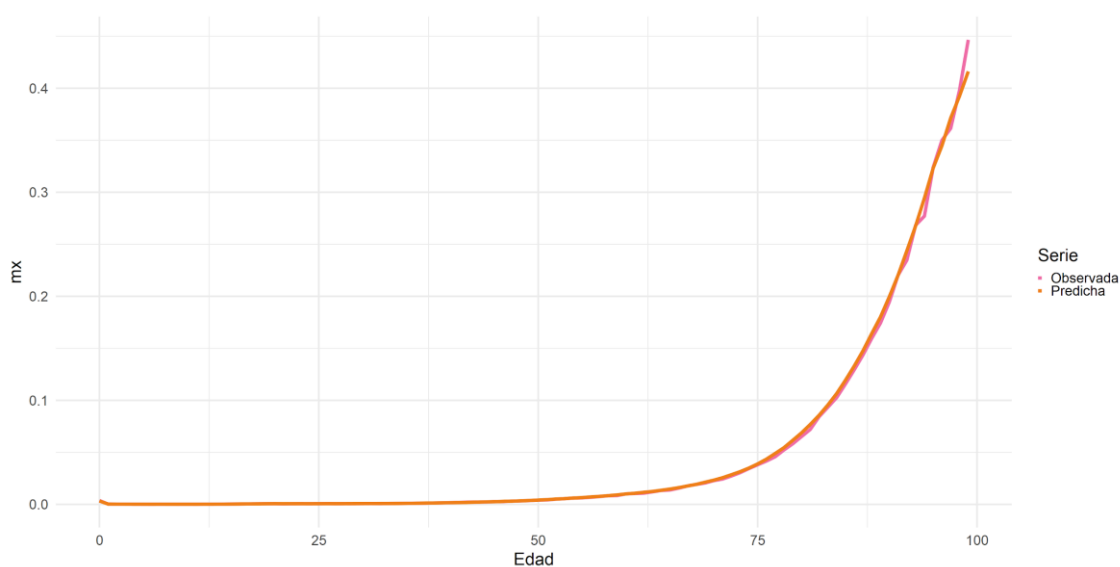


Figura 47. Mortalidad observada y predicha en el año 2007 para el género masculino mediante LSTM.

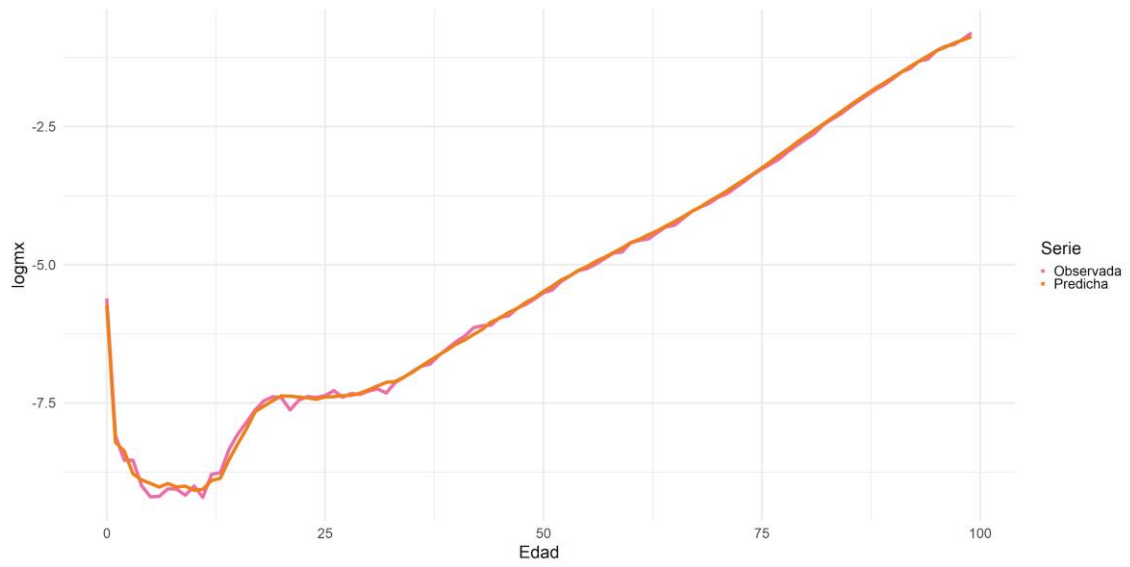


Figura 48. Log-Mortalidad observada y predicha en el año 2007 para el género masculino mediante LSTM.

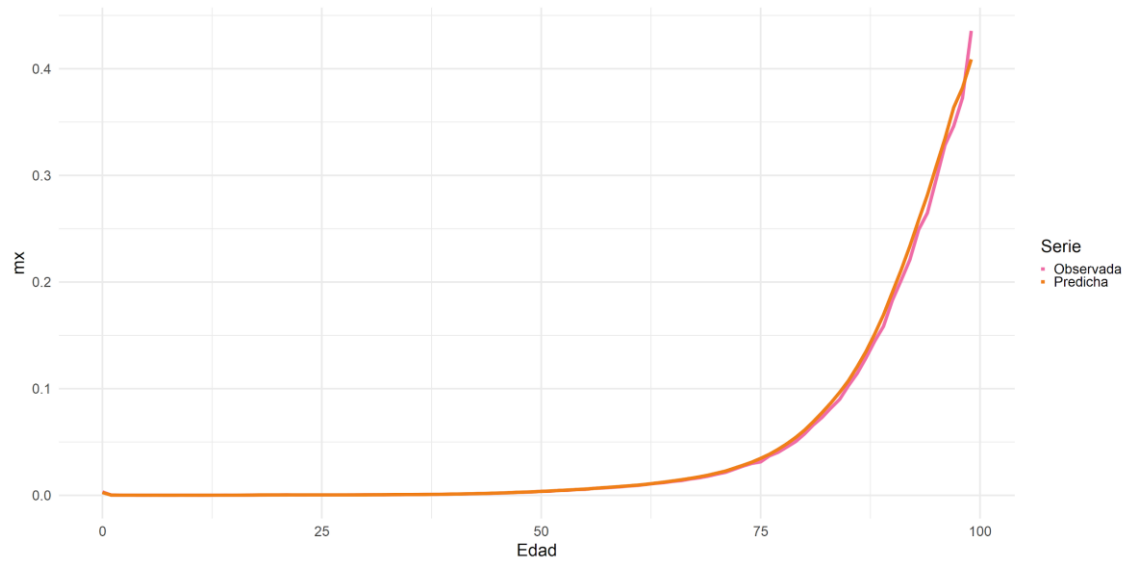


Figura 49. Mortalidad observada y predicha en el año 2011 para el género masculino mediante LSTM.

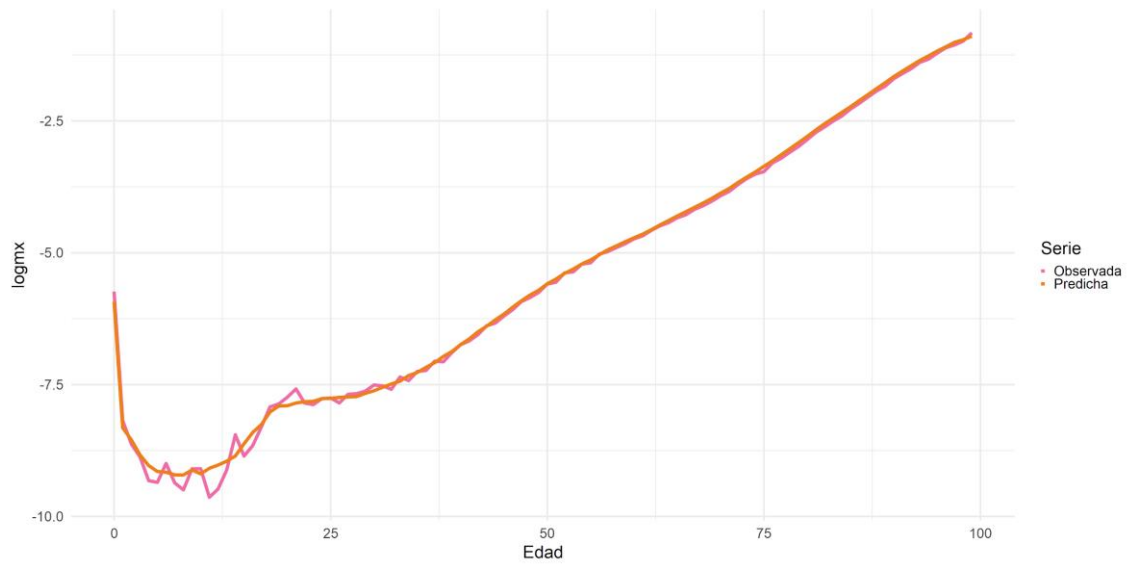


Figura 50. Log-Mortalidad observada y predicha en el año 2011 para el género masculino mediante LSTM.

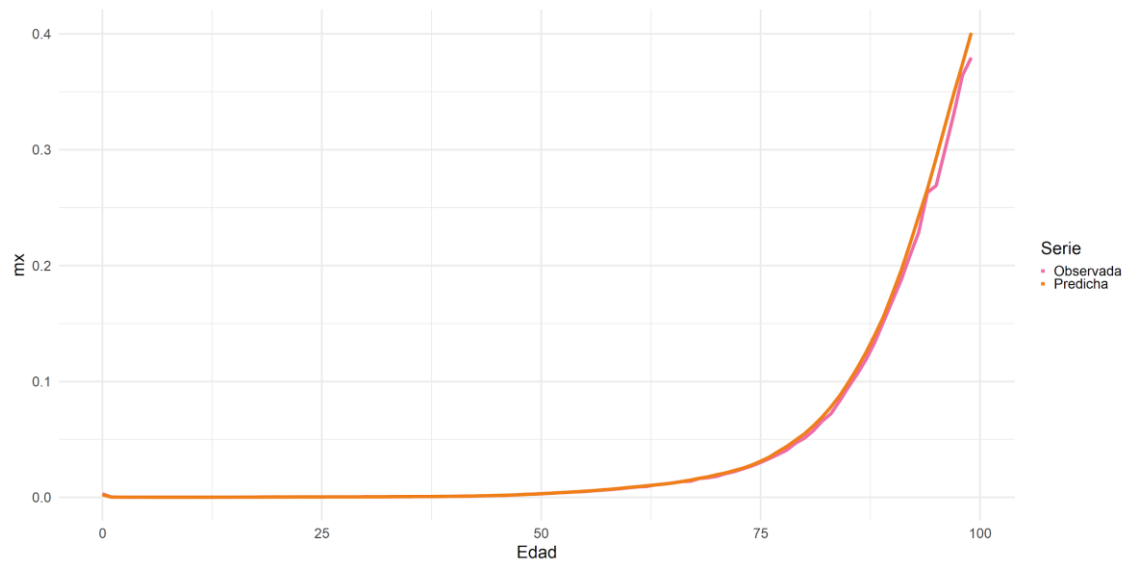


Figura 51. Mortalidad observada y predicha en el año 2016 para el género masculino mediante LSTM.

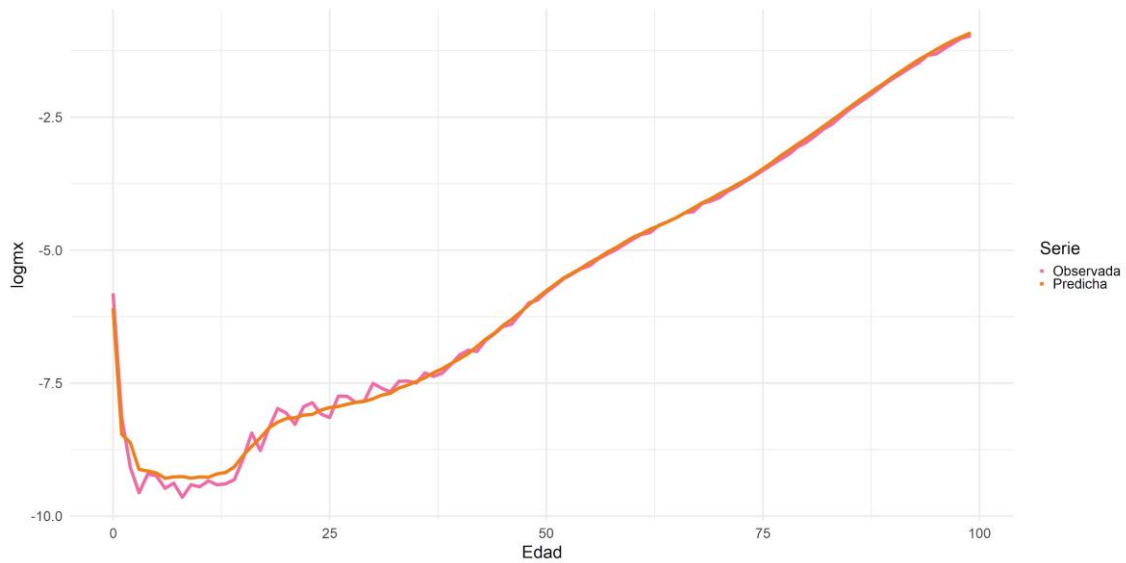


Figura 52. Log-Mortalidad observada y predicha en el año 2016 para el género masculino mediante LSTM.

A diferencia del modelo de Lee Carter, en el cual el rendimiento del modelo empeoraba a la hora de evaluar la población masculina, se observa que mediante el uso de una red neuronal de tipo LSTM se podría solventar el problema. En datos, hablamos de un error porcentual medio absoluto de entre un 6 y un 9% para todas las series, y, además, en medida original, seguiríamos hablando de un error cuadrático medio de mil millonésimas, el cual no sufre grandes alteraciones conforme se aumenta el periodo de proyección.

Tabla 7. Evaluación de la predicción mediante LSTM para el género masculino.

	<b>MSE</b>	<b>MAPE</b>
<b>LSTM 2007 <math>m_{x,t}</math></b>	1.132598e-05	0.0600082
<b>LSTM 2011 <math>m_{x,t}</math></b>	1.050972e-05	0.07340844
<b>LSTM 2016 <math>m_{x,t}</math></b>	8.237059e-06	0.09608338

### 6.3. Análisis conjunto del tramo de edades 80-100

Finalmente, se procede a analizar los tramos de edades más avanzadas, debido a que se ha detectado con anterioridad que son aquellos que más fluctuaciones presentan. Además, este tramo de edades, es interesante para las aseguradoras a la hora de diseñar



sus productos y estrategias financieras, por lo que una predicción más acertada se traduce en menos incertidumbre de forma generalizada. No obstante, toda la información numérica de los análisis realizados en este Capítulo 6 quedan reflejados al completo en los Anexos A y B.

En las Figuras 53, 54 y 55 se representa gráficamente en forma de “zoom” las predicciones para las edades propuestas, en aras de observar en detalle su problemática variabilidad. Para el año 2007, parece que ambos métodos se ajustan considerablemente bien a lo observado, bien es cierto que el modelo de Lee Carter presenta más ruido que el uso de la red LSTM.

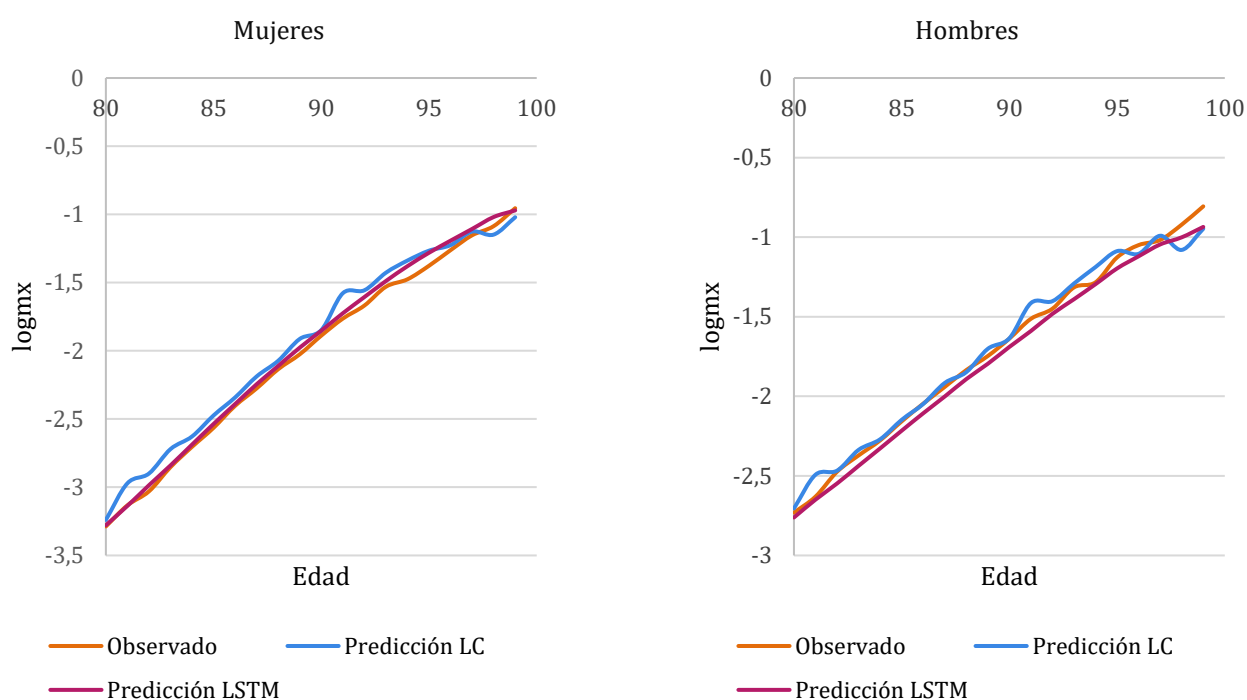


Figura 53. Comparación en el tramo de edades 80-100 para el año 2007.

En el caso del año 2011, se observa como la predicción de la red LSTM sigue la tendencia de lo observado, mientras que el modelo de Lee Carter tiende a sobreestimar la log-mortalidad ligeramente.

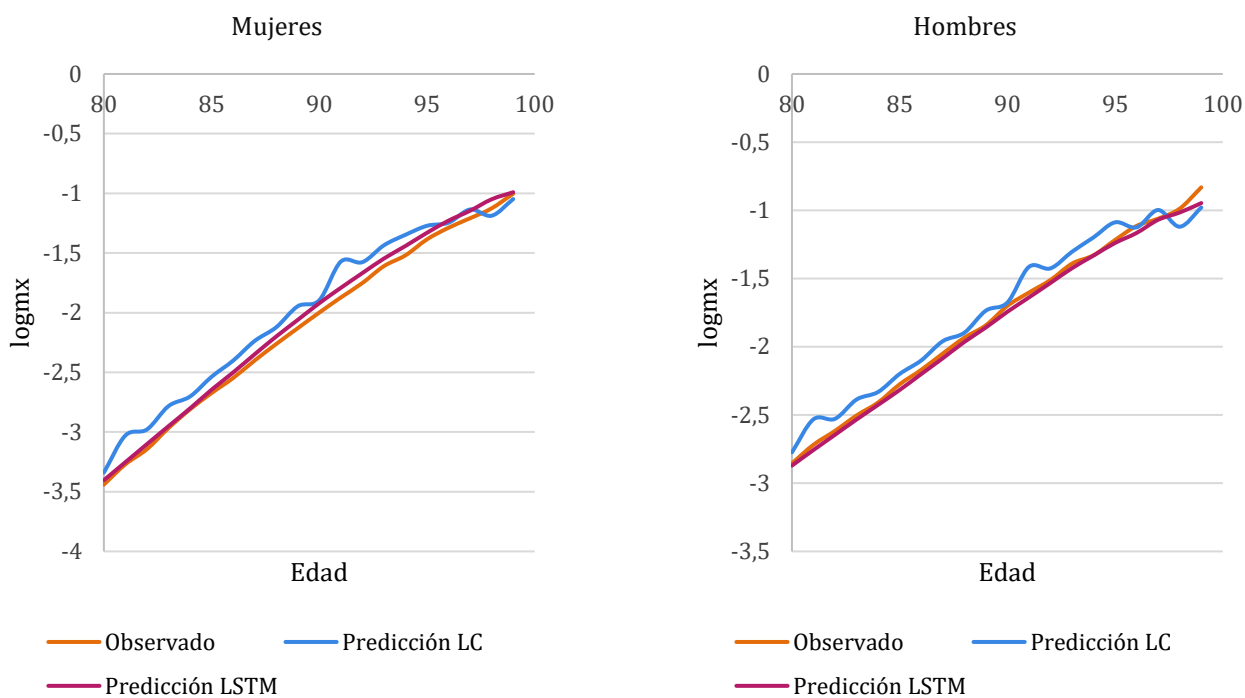


Figura 54. Comparación en el tramo de edades 80-100 para el año 2011.

Para el año 2016, la sobreestimación y el ruido del modelo de Lee Carter aumenta su magnitud de forma considerable, mientras que la red LSTM parece que sigue ajustándose a lo observado dentro de los márgenes establecidos de forma correcta.

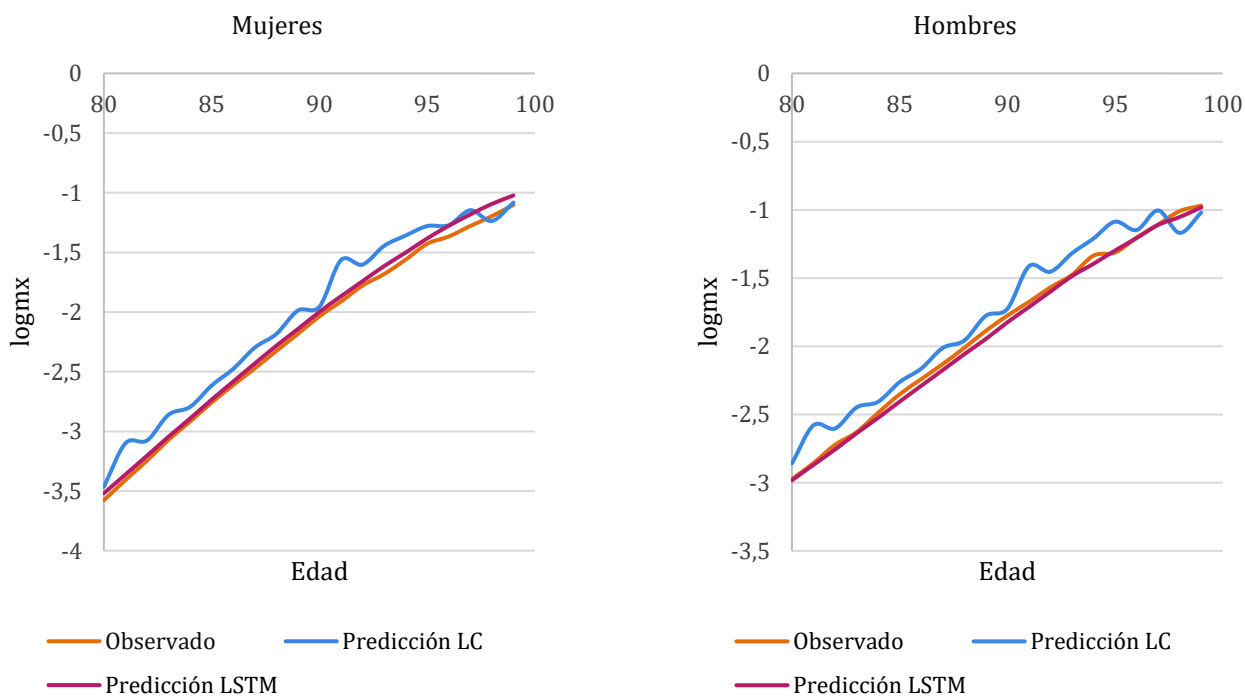


Figura 55. Comparación en el tramo de edades 80-100 para el año 2016.

En las Tablas 8 y 9 podemos encontrar el error porcentual absoluto con respecto a la serie original, calculado para ambos géneros por edad y año con los dos métodos propuestos, el modelo paramétrico de Lee Carter y la red neuronal LSTM. Al final de ellas, aparece reflejado de color gris el error medio porcentual absoluto (MAPE) evaluado específicamente para este tramo de 20 edades.

De forma generalizada se observa que, tanto para el género femenino como para el masculino, el uso de una red neuronal LSTM minimiza el error. Mientras que el modelo de Lee Carter presenta un error medio porcentual absoluto mayor y ascendente conforme aumenta el horizonte temporal para ambos géneros, el error de la red LSTM no sigue un patrón concreto. En el caso de la población femenina, se aprecia que sufre un leve aumento con el tiempo, pero en el caso de la población masculina, se observa que se va minimizando suavizadamente.

El motivo concreto es complicado de hallar, ya que como popularmente se dice, las redes neuronales son una especie de “caja negra” en las que sabemos que inputs son introducidos, que resultados reportan, pero no que está pasando concretamente en las capas ocultas o intermedias. Lo único que se puede afirmar tras los resultados empíricos, es que parece que la red ha aprendido bien con los datos suministrados y mejora la predicción realizada con modelos más tradicionales.

Tabla 8. Zoom en el tramo 80-99 años para el género femenino.

Edad	2007		2011		2016	
	LC	LSTM	LC	LSTM	LC	LSTM
<b>80</b>	1,45%	0,27%	2,92%	1,07%	3,16%	1,63%
<b>81</b>	5,16%	0,13%	7,39%	0,54%	9,03%	1,35%
<b>82</b>	4,31%	1,48%	5,25%	1,34%	5,20%	1,28%
<b>83</b>	4,63%	0,44%	6,20%	0,53%	6,85%	0,87%
<b>84</b>	2,81%	0,61%	3,84%	0,28%	4,21%	0,88%
<b>85</b>	3,54%	1,10%	5,11%	1,05%	5,17%	0,88%
<b>86</b>	2,56%	0,54%	5,73%	1,83%	5,22%	1,16%

<b>87</b>	3,85%	1,34%	6,83%	2,17%	7,04%	1,65%
<b>88</b>	2,90%	1,12%	6,19%	2,77%	6,11%	1,87%
<b>89</b>	5,56%	2,42%	8,62%	3,26%	8,93%	1,84%
<b>90</b>	2,16%	1,90%	5,17%	3,93%	3,97%	1,81%
<b>91</b>	10,56%	2,22%	16,06%	4,32%	18,15%	2,30%
<b>92</b>	6,72%	3,74%	9,98%	4,76%	9,88%	2,05%
<b>93</b>	6,72%	2,74%	10,84%	3,95%	14,07%	3,98%
<b>94</b>	9,32%	6,53%	11,33%	5,15%	12,98%	3,82%
<b>95</b>	8,04%	6,94%	8,19%	3,97%	10,40%	3,18%
<b>96</b>	2,92%	5,51%	3,27%	4,62%	6,93%	6,54%
<b>97</b>	2,35%	4,02%	6,09%	5,23%	10,31%	7,57%
<b>98</b>	5,71%	6,16%	5,49%	6,74%	3,38%	8,56%
<b>99</b>	6,96%	1,71%	4,49%	1,20%	1,76%	7,16%
<b>MAPE</b>	<b>4,91%</b>	<b>2,55%</b>	<b>6,95%</b>	<b>2,93%</b>	<b>7,64%</b>	<b>3,02%</b>

Tabla 9. Zoom en el tramo 80-99 años para el género masculino.

<b>Edad</b>	<b>2007</b>		<b>2011</b>		<b>2016</b>	
	<b>LC</b>	<b>LSTM</b>	<b>LC</b>	<b>LSTM</b>	<b>LC</b>	<b>LSTM</b>
<b>80</b>	0,97%	1,11%	2,84%	0,57%	3,87%	0,27%
<b>81</b>	5,23%	0,72%	6,87%	1,41%	9,70%	0,45%
<b>82</b>	0,27%	2,99%	3,38%	1,09%	4,37%	1,28%
<b>83</b>	1,46%	2,76%	4,67%	1,17%	6,89%	0,32%
<b>84</b>	0,30%	2,21%	3,20%	0,72%	3,24%	1,49%
<b>85</b>	0,44%	2,72%	3,35%	1,82%	3,82%	2,28%
<b>86</b>	0,10%	2,96%	3,19%	1,43%	3,43%	2,19%
<b>87</b>	1,29%	3,10%	4,42%	1,69%	5,54%	2,07%
<b>88</b>	0,70%	3,09%	1,89%	1,61%	2,67%	2,31%
<b>89</b>	2,59%	2,79%	5,81%	0,88%	5,84%	3,15%
<b>90</b>	0,11%	3,19%	1,41%	2,56%	2,78%	2,78%
<b>91</b>	6,54%	5,06%	11,75%	2,24%	15,59%	2,25%

<b>92</b>	3,28%	2,22%	5,65%	1,41%	7,30%	1,99%
<b>93</b>	1,86%	5,64%	6,25%	2,41%	10,71%	0,63%
<b>94</b>	7,46%	0,90%	9,87%	0,16%	9,38%	4,49%
<b>95</b>	3,50%	6,14%	10,63%	1,80%	17,22%	1,11%
<b>96</b>	5,22%	6,68%	0,92%	4,61%	4,69%	0,13%
<b>97</b>	2,60%	2,77%	5,97%	0,77%	9,25%	0,32%
<b>98</b>	17,25%	8,67%	13,43%	2,87%	15,78%	4,26%
<b>99</b>	17,45%	16,04%	17,84%	13,88%	5,26%	1,38%
<b>MAPE</b>	<b>3,93%</b>	<b>4,09%</b>	<b>6,17%</b>	<b>2,25%</b>	<b>7,37%</b>	<b>1,76%</b>

## CAPÍTULO 7. CONCLUSIONES

Los modelos tradicionales de predicción de la mortalidad, de los cuales el modelo de Lee-Carter es su mayor exponente, tienen innegables bondades, llevando a su utilización ser ampliamente aceptada en las Ciencias Actuariales. No obstante, la invariabilidad de los parámetros a lo largo del tiempo y las consecuencias derivadas de ello, dan paso a investigar y desarrollar nuevos métodos dentro de este ámbito que supongan una mejora generalizada de la proyección. Con la introducción y el estudio de las redes neuronales recurrentes de tipo Long Short-Term Memory se permite una mejora de la capacidad predictiva de la mortalidad.

Si mediante el modelo de Lee Carter la predicción generaba errores considerables a partir del grupo de edad de 65 años, este hecho no se observa con el empleo de la red neuronal LSTM. Se observa, que esta tipología de red tiene un rendimiento sobresaliente en todos los tramos de edad y, además, la calidad de la predicción no minora conforme se extiende el horizonte de proyección. Por lo tanto, en el caso de las redes neuronales, el axioma de que a mayor horizonte temporal la predicción empeora no se cumple, a diferencia de lo que acontece en los modelos de mortalidad paramétricos más tradicionales.

De forma concreta, el modelo de Lee Carter tiende a sobrestimar la log-mortalidad en el análisis pormenorizado del tramo de edades 80-100 años, mientras que la predicción de la red LSTM sigue la tendencia de lo observado. En concreto, el modelo de Lee Carter presenta un error medio porcentual absoluto mayor y ascendente conforme aumenta el horizonte temporal para ambos géneros. Sin embargo, el error de la red LSTM no sigue un patrón concreto y es notablemente inferior. Al igual que sucede con las redes neuronales cerebrales, el aprendizaje del entorno permite a las redes neuronales artificiales la adquisición de conocimientos.

Por lo tanto, este tipo de redes de memoria permiten mediante su gran capacidad computacional superar las limitaciones de los modelos tradicionales. Esto es debido a que son redes que clasifican, generan y predicen secuencias. La capacidad computacional de las mismas deriva de poseer una distribución en paralelo y, en definitiva, la red aprende con los datos suministrados y consigue una mejora de la predicción realizada con los modelos tradicionales.

No obstante, la utilización de esta metodología no implica que no haya inconvenientes. La elección de los hiper parámetros que constituyen la configuración básica de la red resulta difícil y tediosa ya que no existe una regla fija sobre ello. Esto supone realizar experimentos a base de prueba y error para encontrar la configuración más óptima, que no implica que lo sea, ya que no es posible probar todas las relaciones de hiper parámetros existentes.

Otro de los inconvenientes que se pueden destacar es que la red no está preparada para predecir situaciones excepcionales o de alto riesgo. En los tiempos que corren, esto hace especial referencia a la pandemia del COVID-19 (SARS-CoV2). Los únicos datos de brotes semejantes son los del SARS-CoV en 2002 y MERS-CoV en 2012, pero no guardan una relación con los datos demográficos anteriores a la fecha de la pandemia, por lo tanto, la red no va a poder predecir estas causas excepcionales.

De cara a futuros trabajos, sería interesante observar si la capacidad predictiva de la red neuronal recurrente de tipo LSTM aumenta para España al añadir nuevas variables relevantes. En este sentido, la introducción de datos demográficos recogidos en países colindantes o con los que se comparta estilos de vida o cultura, como pueden ser Portugal, Italia o Grecia, podrían proporcionar a la red patrones de comportamiento no detectados anteriormente en la dinámica demográfica de los datos originales. Además, también se podría incluir información que aporte un respaldo a las crecientes tasas de longevidad, como el aumento de la calidad de vida, el incremento en los hábitos saludables y los avances médicos presentes a lo largo de los años.

No obstante, como se explicó a lo largo del trabajo, existen otro tipo de redes neuronales recurrentes que también presentan unos resultados sobresalientes. Las redes Gated Recurrent Unit (GRU) son una alternativa un tanto más sencilla a las redes LSTM, pero son utilizadas ampliamente para el estudio y predicción de cadenas secuenciales, como series temporales o síntesis del lenguaje. En muchos casos, no existe una solución óptima para un problema concreto y por ello, ciertos tipos de redes neuronales o configuraciones dentro de una misma, dependiendo de las características del dato bruto, pueden funcionar mejor que otras. El objetivo último es poder realizar unas tablas de mortalidad con una predicción acertada, y el método que se utilice para ello es lo menos relevante siempre y cuando el resultado sea consistente.

## BIBLIOGRAFÍA

- Booth, H., Maindonald, J., & Smith, L. (2002). Applying lee-carter under conditions of variable mortality decline. *Population Studies*, 56(3), 325-336.
- Brownlee, J. (2019, -02-26T18:00:16+00:00). How to use learning curves to diagnose machine learning model performance. Retrieved from <https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/>
- Ciaburro, G. (2017). *Neural networks with R* (1st ed.. ed.) Birmingham: Packt Publishing, Limited.
- Currie, I. D., Durban, M., & Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4), 279-298.
- Falbel, D., Allaire, J. J., Chollet, F., Yuan Tang, Bijl, W. V. D., Studer, M., . . . Keydana, S. (2020). Keras: R interface to 'keras' [computer software]
- Falbel, D., Allaire, J. J., Yuan Tang, Eddelbuettel, D., Golding, N., Kalinowski, T., . . . Tutorials), Google Inc (Examples and. (2020). Tensorflow: R interface to 'TensorFlow' [computer software]
- Fausett, L. (1994). *Fundamentals of neural networks: Architectures, algorithms, and applications* Prentice-Hall, Inc.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* MIT press.
- Gurney, K. (1997). *An introduction to neural networks* CRC press.
- Haykin, S. (2010). *Neural networks and learning machines, 3/E* Pearson Education India.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv Preprint arXiv:1412.6980*,
- Kröse, B., van der Smagt, P., & Smagt, P. (1993). An introduction to neural networks.
- Lee, R. (2000). The lee-carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal*, 4(1), 80-91.
- Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419), 659-671.
- McClelland, J. L., Rumelhart, D. E., & PDP Research Group. (1986). Parallel distributed processing. *Explorations in the Microstructure of Cognition*, 2, 216-271.
- MissingLink. (2018). 7 types of activation functions in neural networks: How to choose? Retrieved from <https://missinglink.ai/guides/neural-network-concepts/7-types->



neural-network-activation-functions-right/

R interface to 'keras'. Retrieved from <https://keras.rstudio.com/>

Ramón, S. (1906). Cajal. *The Nobel Prize in Physiology Or Medicine*,

Rebala, G., Ravi, A. & Churiwala, S. (2019). An introduction to machine learning.

Retrieved from <https://doi.org/10.1007/978-3-030-15729-6>

Renshaw, A. E., & Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3), 556-570.

Rob Hyndman, Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Zhou, Z. (2020). Forecast: Forecasting functions for time series and linear models [computer software]

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

## ANEXO A

### A.1. Análisis en la población femenina para el año 2007

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,003224	-5,736984	0,002755	-5,894261	0,002833	-5,866265
1	0,000333	-8,007289	0,000212	-8,458376	0,000317	-8,058121
2	0,000169	-8,685987	0,000138	-8,889566	0,000182	-8,609011
3	0,000141	-8,863999	0,000105	-9,159461	0,000130	-8,950074
4	0,000103	-9,176536	0,000097	-9,242002	0,000111	-9,108780
5	0,000111	-9,103753	0,000087	-9,347150	0,000095	-9,258448
6	0,000112	-9,094640	0,000075	-9,497352	0,000091	-9,308815
7	0,000081	-9,421248	0,000077	-9,468010	0,000079	-9,444086
8	0,000074	-9,514513	0,000074	-9,512531	0,000084	-9,388798
9	0,000055	-9,805469	0,000079	-9,446988	0,000089	-9,331939
10	0,000105	-9,166192	0,000089	-9,330339	0,000107	-9,143975
11	0,000081	-9,425494	0,000076	-9,483980	0,000099	-9,220083
12	0,000080	-9,427378	0,000090	-9,316968	0,000101	-9,201393
13	0,000122	-9,010210	0,000093	-9,284254	0,000106	-9,156460
14	0,000108	-9,137549	0,000104	-9,175560	0,000107	-9,146910
15	0,000149	-8,814507	0,000144	-8,843428	0,000126	-8,982780
16	0,000143	-8,854065	0,000157	-8,759078	0,000149	-8,811545
17	0,000221	-8,418367	0,000168	-8,689115	0,000165	-8,708782
18	0,000196	-8,535650	0,000192	-8,559353	0,000182	-8,611451
19	0,000164	-8,713091	0,000201	-8,510605	0,000193	-8,550911
20	0,000216	-8,438302	0,000190	-8,565858	0,000188	-8,578459
21	0,000185	-8,593671	0,000192	-8,556097	0,000193	-8,550418
22	0,000214	-8,451138	0,000185	-8,596978	0,000198	-8,529562
23	0,000142	-8,862887	0,000195	-8,542763	0,000193	-8,551445
24	0,000198	-8,525530	0,000202	-8,508566	0,000201	-8,511742
25	0,000169	-8,685642	0,000210	-8,466461	0,000206	-8,488003
26	0,000188	-8,577457	0,000223	-8,410129	0,000218	-8,432631
27	0,000206	-8,485995	0,000232	-8,366779	0,000224	-8,402925
28	0,000211	-8,464606	0,000244	-8,318431	0,000231	-8,371638
29	0,000287	-8,155381	0,000273	-8,205282	0,000240	-8,335795
30	0,000302	-8,105742	0,000289	-8,148985	0,000256	-8,271575
31	0,000296	-8,124669	0,000336	-7,998840	0,000287	-8,156485
32	0,000305	-8,094809	0,000338	-7,993578	0,000321	-8,043285
33	0,000345	-7,972076	0,000379	-7,878325	0,000362	-7,922748
34	0,000381	-7,873476	0,000391	-7,846821	0,000404	-7,814579
35	0,000382	-7,868968	0,000429	-7,754581	0,000445	-7,717012
36	0,000494	-7,613013	0,000470	-7,663160	0,000502	-7,596039
37	0,000539	-7,525745	0,000507	-7,586624	0,000572	-7,466916
38	0,000611	-7,400860	0,000532	-7,538277	0,000640	-7,354249

39	0,000737	-7,212613	0,000624	-7,378599	0,000722	-7,233017
40	0,000716	-7,241801	0,000630	-7,369123	0,000755	-7,188456
41	0,000787	-7,147160	0,000764	-7,176871	0,000844	-7,077508
42	0,000999	-6,909107	0,000755	-7,188423	0,000904	-7,008956
43	0,000903	-7,009401	0,000896	-7,018045	0,000991	-6,917167
44	0,001107	-6,806496	0,000982	-6,926098	0,001085	-6,825973
45	0,001264	-6,673522	0,001031	-6,876769	0,001196	-6,728609
46	0,001300	-6,645036	0,001132	-6,783619	0,001296	-6,648510
47	0,001497	-6,503972	0,001296	-6,648249	0,001371	-6,592084
48	0,001541	-6,475098	0,001345	-6,611091	0,001481	-6,514776
49	0,001653	-6,405108	0,001519	-6,489830	0,001576	-6,452669
50	0,001894	-6,269209	0,001538	-6,477342	0,001689	-6,383440
51	0,002016	-6,206397	0,001821	-6,308303	0,001833	-6,301790
52	0,002040	-6,194825	0,001818	-6,310191	0,001974	-6,227608
53	0,002326	-6,063419	0,002055	-6,187259	0,002112	-6,160145
54	0,002383	-6,039553	0,002128	-6,152579	0,002231	-6,105310
55	0,002511	-5,987170	0,002354	-6,051584	0,002407	-6,029541
56	0,002517	-5,984609	0,002489	-5,995899	0,002510	-5,987594
57	0,002784	-5,883801	0,002818	-5,871647	0,002754	-5,894773
58	0,003196	-5,745879	0,002893	-5,845446	0,002975	-5,817611
59	0,003283	-5,718854	0,003366	-5,693898	0,003243	-5,731129
60	0,003762	-5,582817	0,003384	-5,688630	0,003459	-5,666652
61	0,003866	-5,555615	0,004070	-5,504012	0,003772	-5,580100
62	0,004381	-5,430506	0,004180	-5,477475	0,004126	-5,490536
63	0,004457	-5,413184	0,004678	-5,364837	0,004500	-5,403783
64	0,005225	-5,254394	0,005119	-5,274857	0,004986	-5,301184
65	0,005495	-5,204000	0,005532	-5,197287	0,005455	-5,211249
66	0,005954	-5,123617	0,006334	-5,061861	0,006022	-5,112375
67	0,006710	-5,004187	0,007017	-4,959381	0,006672	-5,009818
68	0,007672	-4,870165	0,007751	-4,859970	0,007386	-4,908192
69	0,008847	-4,727646	0,009447	-4,662072	0,008395	-4,780149
70	0,009431	-4,663790	0,009550	-4,651228	0,009424	-4,664515
71	0,010596	-4,547311	0,011989	-4,423798	0,010641	-4,543081
72	0,012170	-4,408761	0,012608	-4,373458	0,012109	-4,413839
73	0,013262	-4,322825	0,014834	-4,210832	0,013772	-4,285137
74	0,016100	-4,128957	0,017013	-4,073770	0,015935	-4,139266
75	0,017902	-4,022819	0,019483	-3,938193	0,018101	-4,011786
76	0,020621	-3,881437	0,022770	-3,782322	0,020893	-3,868349
77	0,023824	-3,737076	0,026673	-3,624110	0,024205	-3,721190
78	0,027670	-3,587391	0,029811	-3,512873	0,028019	-3,574882
79	0,031753	-3,449773	0,037513	-3,283077	0,032526	-3,425729
80	0,037352	-3,287377	0,039175	-3,239721	0,037683	-3,278552
81	0,043609	-3,132493	0,051254	-2,970957	0,043437	-3,136453
82	0,048304	-3,030238	0,055049	-2,899525	0,050527	-2,985246

83	0,057705	-2,852406	0,065861	-2,720216	0,058429	-2,839936
84	0,066959	-2,703669	0,072243	-2,627724	0,068073	-2,687168
85	0,076983	-2,564165	0,084290	-2,473490	0,079189	-2,535921
86	0,090458	-2,402872	0,096191	-2,341419	0,091636	-2,389928
87	0,102787	-2,275095	0,112196	-2,187512	0,105972	-2,244581
88	0,118295	-2,134577	0,125850	-2,072662	0,121158	-2,110658
89	0,131944	-2,025376	0,147668	-1,912791	0,138559	-1,976461
90	0,151319	-1,888367	0,157609	-1,847638	0,156842	-1,852514
91	0,171437	-1,763539	0,206526	-1,577327	0,178279	-1,724404
92	0,188582	-1,668221	0,210970	-1,556041	0,200732	-1,605782
93	0,216515	-1,530097	0,239971	-1,427237	0,225787	-1,488162
94	0,228656	-1,475538	0,262354	-1,338060	0,251765	-1,379257
95	0,252483	-1,376411	0,282018	-1,265783	0,277787	-1,280901
96	0,283121	-1,261880	0,293751	-1,225023	0,303500	-1,192374
97	0,315714	-1,152917	0,324403	-1,125768	0,330693	-1,106565
98	0,337568	-1,085987	0,317283	-1,147961	0,360923	-1,019091
99	0,385084	-0,954293	0,360354	-1,020668	0,378856	-0,970600

## A.2. Análisis en la población femenina para el año 2011

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,002876	-5,85132	0,002112	-6,16007	0,00266	-5,9294
1	0,000277	-8,1916	0,000159	-8,74529	0,000325	-8,03025
2	0,000114	-9,08047	0,000105	-9,16238	0,000166	-8,70102
3	0,000105	-9,16211	8,12E-05	-9,4182	0,000121	-9,01725
4	9,11E-05	-9,30347	7,64E-05	-9,47928	0,000108	-9,13301
5	7,94E-05	-9,44162	6,97E-05	-9,57194	0,000101	-9,20166
6	9,75E-05	-9,23529	5,97E-05	-9,72631	7,83E-05	-9,45509
7	5,63E-05	-9,78459	6,25E-05	-9,68114	7,23E-05	-9,53485
8	5,29E-05	-9,84767	6,01E-05	-9,71869	7,25E-05	-9,53252
9	5,91E-05	-9,73639	6,56E-05	-9,6317	7,2E-05	-9,53854
10	8,22E-05	-9,40614	7,48E-05	-9,50034	8,19E-05	-9,41024
11	7,85E-05	-9,45195	6,34E-05	-9,66634	9,08E-05	-9,3063
12	8,1E-05	-9,42097	7,55E-05	-9,49175	9,2E-05	-9,29326
13	0,000121	-9,01965	7,86E-05	-9,45067	9,25E-05	-9,28818
14	0,000129	-8,95376	8,76E-05	-9,34318	8,9E-05	-9,32633
15	0,000126	-8,97629	0,000126	-8,97673	0,0001	-9,20548
16	0,000112	-9,09779	0,000137	-8,89335	0,000114	-9,0811
17	0,000203	-8,50392	0,000147	-8,82207	0,000131	-8,93785
18	0,000148	-8,81787	0,000168	-8,68868	0,000154	-8,78038
19	0,000225	-8,39918	0,000178	-8,6364	0,000148	-8,81512
20	0,000147	-8,82406	0,000165	-8,70832	0,000154	-8,78029
21	0,000125	-8,98348	0,000167	-8,69569	0,000148	-8,81745

22	0,000165	-8,711	0,000157	-8,75812	0,000153	-8,78216
23	0,000143	-8,84992	0,000166	-8,70083	0,000153	-8,78456
24	0,000151	-8,79499	0,000172	-8,66759	0,000154	-8,78155
25	0,000204	-8,49758	0,000179	-8,62609	0,000161	-8,73668
26	0,000182	-8,60961	0,000189	-8,57151	0,000163	-8,7235
27	0,000191	-8,56096	0,000198	-8,5287	0,000176	-8,64759
28	0,000194	-8,54786	0,000208	-8,47955	0,000188	-8,58052
29	0,000208	-8,47722	0,000236	-8,3518	0,000208	-8,47632
30	0,000262	-8,24803	0,000248	-8,30216	0,000223	-8,40724
31	0,000283	-8,17023	0,000296	-8,12584	0,000242	-8,32582
32	0,000282	-8,17263	0,000292	-8,13776	0,000257	-8,26468
33	0,000257	-8,26714	0,000331	-8,01209	0,000291	-8,14355
34	0,000329	-8,0192	0,000341	-7,98221	0,000323	-8,03742
35	0,000324	-8,03435	0,000375	-7,88806	0,000367	-7,91105
36	0,000417	-7,78132	0,000413	-7,79266	0,000406	-7,80946
37	0,000442	-7,72397	0,000448	-7,71041	0,000439	-7,73181
38	0,000519	-7,56375	0,000466	-7,67097	0,000491	-7,61999
39	0,000587	-7,44074	0,000558	-7,49079	0,000551	-7,50437
40	0,00061	-7,40219	0,000552	-7,50237	0,000605	-7,41105
41	0,000661	-7,3222	0,000693	-7,2743	0,000691	-7,2778
42	0,000819	-7,1075	0,000666	-7,3138	0,000777	-7,16004
43	0,000965	-6,94389	0,000808	-7,12138	0,000858	-7,06079
44	0,001007	-6,90094	0,000887	-7,02796	0,000961	-6,94735
45	0,001	-6,90804	0,000924	-6,98638	0,001086	-6,82569
46	0,001148	-6,76962	0,001019	-6,88863	0,001181	-6,7414
47	0,001341	-6,61465	0,001178	-6,74384	0,001273	-6,66667
48	0,001436	-6,54571	0,001207	-6,71951	0,001387	-6,58039
49	0,001683	-6,38743	0,001383	-6,58315	0,00152	-6,48925
50	0,001656	-6,4032	0,001371	-6,59254	0,001625	-6,42228
51	0,002116	-6,1583	0,001671	-6,39448	0,001775	-6,3341
52	0,00217	-6,13293	0,001624	-6,42281	0,001916	-6,25743
53	0,002239	-6,10172	0,001863	-6,28583	0,002128	-6,1525
54	0,002231	-6,10514	0,001908	-6,26166	0,002259	-6,09278
55	0,002299	-6,07524	0,002127	-6,15298	0,002439	-6,01611
56	0,002831	-5,86725	0,00223	-6,10577	0,00262	-5,94452
57	0,002967	-5,8201	0,002551	-5,97138	0,002797	-5,8791
58	0,002983	-5,81498	0,002577	-5,96123	0,003028	-5,80002
59	0,003566	-5,63622	0,003045	-5,79423	0,003203	-5,74353
60	0,003584	-5,63138	0,002984	-5,81452	0,003345	-5,70022
61	0,003711	-5,59641	0,003702	-5,59883	0,003596	-5,6278
62	0,003984	-5,52539	0,003711	-5,59655	0,00384	-5,56222
63	0,004366	-5,4339	0,004174	-5,479	0,004086	-5,50028
64	0,00446	-5,41253	0,004555	-5,39157	0,004506	-5,40224
65	0,004996	-5,29902	0,004884	-5,32186	0,004921	-5,31426

66	0,005209	-5,25732	0,005649	-5,17624	0,005417	-5,21823
67	0,005977	-5,11976	0,006238	-5,07707	0,005982	-5,11894
68	0,006854	-4,98293	0,006869	-4,98075	0,006607	-5,01956
69	0,007833	-4,84942	0,008547	-4,76223	0,007426	-4,9028
70	0,008317	-4,78943	0,008377	-4,78232	0,008363	-4,78394
71	0,009155	-4,69351	0,010864	-4,52232	0,009426	-4,66424
72	0,010576	-4,54918	0,011134	-4,49778	0,010795	-4,52863
73	0,01194	-4,42788	0,013256	-4,32334	0,01222	-4,40466
74	0,013743	-4,28725	0,015191	-4,18704	0,013801	-4,28299
75	0,015874	-4,14307	0,017442	-4,04887	0,015872	-4,14318
76	0,017621	-4,03866	0,020507	-3,887	0,018079	-4,01298
77	0,020888	-3,86857	0,024211	-3,72096	0,020955	-3,8654
78	0,023762	-3,73967	0,026796	-3,61952	0,024323	-3,71632
79	0,028299	-3,56491	0,03471	-3,36073	0,028497	-3,55795
80	0,032065	-3,44	0,035454	-3,33953	0,033263	-3,40332
81	0,038005	-3,27003	0,04839	-3,02846	0,038687	-3,25226
82	0,043146	-3,14317	0,050884	-2,97821	0,044996	-3,10118
83	0,05148	-2,96656	0,06188	-2,78256	0,052303	-2,9507
84	0,060211	-2,8099	0,06707	-2,70202	0,060693	-2,80193
85	0,069033	-2,67318	0,079131	-2,53665	0,070988	-2,64524
86	0,078295	-2,54727	0,090592	-2,40139	0,082022	-2,50077
87	0,090685	-2,40037	0,106828	-2,23654	0,095537	-2,34824
88	0,104119	-2,26222	0,119778	-2,12211	0,110861	-2,19948
89	0,118896	-2,1295	0,142844	-1,946	0,127439	-2,06011
90	0,13551	-1,99871	0,150268	-1,89534	0,146573	-1,92023
91	0,153752	-1,87242	0,207689	-1,57171	0,166694	-1,7916
92	0,173525	-1,75143	0,206674	-1,57661	0,188612	-1,66806
93	0,19993	-1,60979	0,23804	-1,43532	0,213048	-1,54624
94	0,218865	-1,5193	0,259964	-1,34721	0,236684	-1,44103
95	0,250252	-1,38529	0,280308	-1,27187	0,264402	-1,33028
96	0,275803	-1,28807	0,287662	-1,24597	0,2927	-1,22861
97	0,298589	-1,20869	0,3214	-1,13507	0,318058	-1,14552
98	0,324447	-1,12563	0,305002	-1,18744	0,350016	-1,04978
99	0,366801	-1,00294	0,350637	-1,048	0,371257	-0,99086

### A.3. Análisis en la población femenina para el año 2016.

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,002262	-6,09153	0,001515	-6,49233	0,002353	-6,05211
1	0,000218	-8,43254	0,000111	-9,10393	0,000232	-8,36831
2	0,000144	-8,84414	7,46E-05	-9,50339	0,00013	-8,94869
3	7,02E-05	-9,56433	5,88E-05	-9,74163	8,71E-05	-9,34824

4	8,87E-05	-9,33052	5,68E-05	-9,77589	8,03E-05	-9,43013
5	5,17E-05	-9,87017	5,26E-05	-9,85294	8,4E-05	-9,38506
6	8,18E-05	-9,41153	4,48E-05	-10,0125	8,09E-05	-9,42279
7	1,23E-05	-11,3045	4,78E-05	-9,94755	8,21E-05	-9,40728
8	6,92E-05	-9,5791	4,65E-05	-9,97639	8,39E-05	-9,3863
9	7,5E-05	-9,49839	5,21E-05	-9,86259	8,24E-05	-9,40344
10	5,03E-05	-9,89718	6,05E-05	-9,71284	7,91E-05	-9,44421
11	5,52E-05	-9,80379	5,05E-05	-9,89428	8,27E-05	-9,40084
12	6,94E-05	-9,57541	6,07E-05	-9,71022	8,6E-05	-9,3608
13	6,17E-05	-9,69302	6,39E-05	-9,65868	8,45E-05	-9,37868
14	7,72E-05	-9,46925	7,1E-05	-9,5527	9,1E-05	-9,30512
15	0,000132	-8,9316	0,000107	-9,14336	8,81E-05	-9,3376
16	0,000101	-9,19743	0,000116	-9,06119	9,71E-05	-9,23951
17	0,000104	-9,16714	0,000125	-8,98825	0,000106	-9,15293
18	0,00011	-9,11257	0,000143	-8,85034	0,000115	-9,07118
19	0,00016	-8,74123	0,000152	-8,79364	0,00012	-9,02717
20	0,000142	-8,85782	0,000138	-8,8864	0,000125	-8,9876
21	0,000127	-8,9697	0,000141	-8,87017	0,000131	-8,94182
22	0,000154	-8,77808	0,000129	-8,95955	0,000128	-8,96347
23	0,000125	-8,98853	0,000137	-8,89842	0,000141	-8,86782
24	0,000148	-8,81509	0,000141	-8,86636	0,000137	-8,89689
25	0,000159	-8,74962	0,000147	-8,82563	0,000141	-8,86697
26	0,00015	-8,80261	0,000155	-8,77323	0,000144	-8,84772
27	0,000162	-8,7271	0,000161	-8,73111	0,000153	-8,78696
28	0,000154	-8,77699	0,00017	-8,68095	0,000164	-8,7146
29	0,000245	-8,31431	0,000196	-8,53495	0,00017	-8,67717
30	0,000236	-8,35248	0,000205	-8,49363	0,000183	-8,60833
31	0,000217	-8,43472	0,000252	-8,28459	0,000195	-8,54051
32	0,000203	-8,50391	0,000244	-8,31799	0,000204	-8,49547
33	0,000234	-8,36036	0,00028	-8,1793	0,000227	-8,39138
34	0,000287	-8,15769	0,000288	-8,15145	0,000264	-8,23836
35	0,000338	-7,99249	0,000318	-8,0549	0,000293	-8,13703
36	0,000321	-8,04386	0,000351	-7,95455	0,000334	-8,00459
37	0,000417	-7,78266	0,000384	-7,86515	0,000368	-7,90769
38	0,000399	-7,8257	0,000395	-7,83683	0,000411	-7,7958
39	0,000493	-7,61489	0,000485	-7,63102	0,000445	-7,71724
40	0,000548	-7,50962	0,000467	-7,66892	0,000521	-7,5597
41	0,000623	-7,38086	0,000614	-7,39608	0,000589	-7,43757
42	0,000629	-7,37067	0,00057	-7,47053	0,000673	-7,3038
43	0,000653	-7,33407	0,00071	-7,25055	0,00076	-7,18247
44	0,000786	-7,14852	0,000781	-7,15529	0,000856	-7,063
45	0,00093	-6,98034	0,000806	-7,12339	0,000939	-6,97083
46	0,001022	-6,8863	0,000894	-7,01988	0,00101	-6,89748
47	0,001137	-6,77958	0,001045	-6,86333	0,001161	-6,75879

48	0,001334	-6,61993	0,001054	-6,85504	0,001249	-6,68578
49	0,001294	-6,6501	0,001231	-6,69981	0,001384	-6,58307
50	0,001562	-6,4616	0,001187	-6,73654	0,001521	-6,48809
51	0,001791	-6,32524	0,0015	-6,5022	0,001676	-6,39134
52	0,001981	-6,22427	0,001411	-6,56358	0,001845	-6,29545
53	0,002162	-6,13684	0,001647	-6,40904	0,002047	-6,19123
54	0,002187	-6,12526	0,001665	-6,398	0,002225	-6,10786
55	0,002417	-6,02519	0,001874	-6,27973	0,002435	-6,01783
56	0,002703	-5,91341	0,001944	-6,24311	0,002664	-5,92787
57	0,002919	-5,83668	0,002252	-6,09604	0,002901	-5,84269
58	0,003067	-5,78695	0,00223	-6,10597	0,003125	-5,76824
59	0,003196	-5,74599	0,002686	-5,91963	0,003297	-5,71465
60	0,003548	-5,64133	0,002549	-5,97187	0,003509	-5,65233
61	0,003805	-5,57144	0,003288	-5,71736	0,003698	-5,60008
62	0,003969	-5,52937	0,003197	-5,74539	0,003925	-5,54043
63	0,004278	-5,45423	0,003618	-5,6217	0,004166	-5,48071
64	0,004887	-5,32116	0,003937	-5,53746	0,004528	-5,39748
65	0,00509	-5,28056	0,004179	-5,47758	0,004886	-5,32136
66	0,005467	-5,20898	0,004897	-5,31921	0,005289	-5,24219
67	0,006025	-5,11178	0,005385	-5,22417	0,005804	-5,14915
68	0,006358	-5,05811	0,005906	-5,13172	0,006333	-5,06205
69	0,007196	-4,93426	0,007541	-4,88742	0,007072	-4,95159
70	0,007668	-4,87066	0,00711	-4,94619	0,007767	-4,8579
71	0,008942	-4,71695	0,009605	-4,64548	0,008682	-4,74655
72	0,009793	-4,62605	0,009531	-4,65319	0,009662	-4,63955
73	0,010704	-4,53718	0,011516	-4,46398	0,010919	-4,51726
74	0,012718	-4,36472	0,013186	-4,32863	0,012316	-4,39688
75	0,014404	-4,24027	0,015188	-4,18722	0,014162	-4,25717
76	0,015799	-4,14782	0,017992	-4,01785	0,016252	-4,11951
77	0,01881	-3,97335	0,02145	-3,84202	0,018832	-3,97222
78	0,021227	-3,85249	0,023452	-3,75282	0,021774	-3,82705
79	0,025153	-3,68278	0,031499	-3,45779	0,025395	-3,6732
80	0,027952	-3,57726	0,031295	-3,4643	0,029632	-3,5189
81	0,033109	-3,40794	0,045034	-3,10033	0,034672	-3,36183
82	0,038954	-3,24536	0,046118	-3,07656	0,040604	-3,20388
83	0,046378	-3,07093	0,057241	-2,86049	0,04763	-3,0443
84	0,054064	-2,91758	0,061122	-2,79489	0,055468	-2,89195
85	0,063407	-2,75817	0,073123	-2,61561	0,064971	-2,73381
86	0,07334	-2,61265	0,084049	-2,47636	0,075601	-2,58229
87	0,08443	-2,47183	0,100478	-2,29782	0,087934	-2,43117
88	0,097672	-2,32614	0,112599	-2,18393	0,102011	-2,28268
89	0,112762	-2,18247	0,137035	-1,98752	0,117388	-2,14227
90	0,130579	-2,03577	0,14157	-1,95496	0,135482	-1,99892
91	0,147827	-1,91171	0,209151	-1,5647	0,154466	-1,86778



92	0,168967	-1,77805	0,201427	-1,60233	0,175244	-1,74158
93	0,186003	-1,68199	0,235648	-1,44542	0,198883	-1,61504
94	0,209875	-1,56124	0,257006	-1,35866	0,222768	-1,50162
95	0,239778	-1,42804	0,278184	-1,27947	0,250926	-1,3826
96	0,254901	-1,36688	0,280229	-1,27215	0,278719	-1,27755
97	0,278435	-1,27857	0,317684	-1,1467	0,306724	-1,18181
98	0,302314	-1,19629	0,290318	-1,23678	0,334894	-1,09394
99	0,332348	-1,10157	0,338859	-1,08217	0,359632	-1,02268

## ANEXO B

### B.1. Análisis en la población masculina para el año 2007

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,003664	-5,60916	0,003037	-5,79681	0,002665	-5,92737
1	0,000306	-8,09246	0,000215	-8,44459	0,000294	-8,13289
2	0,000196	-8,5357	0,000153	-8,7828	0,000246	-8,31135
3	0,000196	-8,53582	0,000136	-8,90577	0,000166	-8,70584
4	0,000124	-8,99912	0,000118	-9,04821	0,00014	-8,87399
5	0,000101	-9,19777	0,000111	-9,10541	0,000136	-8,90205
6	0,000102	-9,191	0,000117	-9,05216	0,000124	-8,99753
7	0,000117	-9,05221	0,000109	-9,12383	0,000127	-8,97248
8	0,000116	-9,05919	0,00011	-9,11347	0,000125	-8,98855
9	0,000104	-9,16858	9,92E-05	-9,21884	0,00012	-9,0281
10	0,000123	-9,00356	0,000103	-9,17668	0,000117	-9,05455
11	0,0001	-9,20554	0,000106	-9,14998	0,000115	-9,06776
12	0,000152	-8,79069	0,00012	-9,02402	0,000142	-8,86065
13	0,000157	-8,76115	0,000151	-8,79641	0,000148	-8,81586
14	0,000239	-8,34032	0,00019	-8,5671	0,000193	-8,55163
15	0,000315	-8,06271	0,000268	-8,22571	0,000276	-8,19416
16	0,00039	-7,84947	0,000411	-7,79802	0,000352	-7,95314
17	0,000488	-7,62574	0,00053	-7,54269	0,000476	-7,65063
18	0,000576	-7,46006	0,000609	-7,40368	0,000519	-7,56408
19	0,000619	-7,38686	0,000686	-7,28503	0,000577	-7,45832
20	0,000608	-7,40523	0,000729	-7,22349	0,000632	-7,36663
21	0,000488	-7,6259	0,000752	-7,19307	0,00062	-7,385
22	0,000582	-7,44928	0,000761	-7,18107	0,000621	-7,38401
23	0,000621	-7,38374	0,000765	-7,175	0,000622	-7,38247
24	0,000611	-7,4005	0,000743	-7,20422	0,000602	-7,41606
25	0,000632	-7,36678	0,000779	-7,1569	0,00062	-7,38515
26	0,000691	-7,27716	0,000778	-7,15877	0,000607	-7,40709
27	0,000613	-7,39672	0,000823	-7,10214	0,000635	-7,3616
28	0,000656	-7,32964	0,000834	-7,08885	0,00065	-7,33923
29	0,000645	-7,34578	0,000938	-6,97142	0,000666	-7,31421
30	0,000686	-7,28393	0,00096	-6,94817	0,000692	-7,27623
31	0,000713	-7,24597	0,001058	-6,85184	0,000733	-7,2182
32	0,000662	-7,32042	0,00106	-6,84903	0,000794	-7,13795
33	0,000798	-7,13379	0,00113	-6,78598	0,000833	-7,09042
34	0,000878	-7,03798	0,001182	-6,74073	0,000873	-7,04385
35	0,000977	-6,93065	0,001241	-6,69167	0,000955	-6,95359
36	0,001074	-6,83616	0,001305	-6,64159	0,001068	-6,84243
37	0,001115	-6,79902	0,001394	-6,57548	0,001197	-6,72817
38	0,001302	-6,64385	0,001471	-6,5221	0,001318	-6,63144

39	0,001477	-6,51773	0,001666	-6,39709	0,001414	-6,56105
40	0,001678	-6,39045	0,001675	-6,39208	0,001579	-6,45116
41	0,001857	-6,28887	0,001905	-6,26346	0,001734	-6,3573
42	0,002157	-6,13917	0,001923	-6,25379	0,001916	-6,25729
43	0,002237	-6,10249	0,002208	-6,11578	0,002047	-6,19122
44	0,00226	-6,09221	0,002377	-6,0419	0,002338	-6,05835
45	0,002602	-5,95148	0,002586	-5,95766	0,002543	-5,97428
46	0,002669	-5,92616	0,002861	-5,85659	0,002805	-5,87625
47	0,003079	-5,78326	0,003161	-5,75679	0,003018	-5,80328
48	0,003294	-5,71568	0,00338	-5,6899	0,0034	-5,68412
49	0,003636	-5,61689	0,003825	-5,56627	0,003689	-5,6024
50	0,004057	-5,50732	0,003991	-5,52364	0,004127	-5,49032
51	0,004266	-5,45699	0,004609	-5,37969	0,004474	-5,40939
52	0,004967	-5,30502	0,004766	-5,34615	0,005024	-5,29353
53	0,005488	-5,20511	0,00533	-5,23438	0,005397	-5,2219
54	0,006054	-5,10705	0,005796	-5,15053	0,005927	-5,12825
55	0,006286	-5,06939	0,006271	-5,07188	0,006355	-5,05847
56	0,006815	-4,98857	0,006859	-4,98224	0,006951	-4,96892
57	0,007534	-4,88839	0,007584	-4,88177	0,007488	-4,89443
58	0,008318	-4,78939	0,008031	-4,82448	0,008153	-4,80936
59	0,008488	-4,76914	0,009161	-4,69278	0,008706	-4,74378
60	0,010062	-4,59897	0,009628	-4,64306	0,009548	-4,65141
61	0,010475	-4,55874	0,01112	-4,49897	0,010187	-4,58664
62	0,010739	-4,53391	0,011329	-4,48038	0,011009	-4,50904
63	0,012004	-4,42252	0,012465	-4,38481	0,0118	-4,43962
64	0,013366	-4,31504	0,013878	-4,27746	0,012828	-4,35609
65	0,013772	-4,2851	0,014953	-4,20286	0,013958	-4,27171
66	0,015586	-4,16138	0,016771	-4,08808	0,015253	-4,18301
67	0,017682	-4,03519	0,018307	-4,00048	0,016718	-4,09127
68	0,019168	-3,95452	0,019879	-3,91808	0,018116	-4,01096
69	0,020513	-3,8867	0,022894	-3,77688	0,019959	-3,91406
70	0,022805	-3,78077	0,023524	-3,74975	0,021904	-3,8211
71	0,024333	-3,71592	0,027973	-3,57652	0,02397	-3,73096
72	0,027228	-3,60351	0,028813	-3,54693	0,02665	-3,62497
73	0,030457	-3,49144	0,032494	-3,4267	0,029438	-3,52546
74	0,034389	-3,37001	0,035853	-3,32834	0,032665	-3,42144
75	0,038023	-3,26955	0,03967	-3,22717	0,036346	-3,31466
76	0,041631	-3,17892	0,04458	-3,11046	0,040531	-3,20569
77	0,045771	-3,0841	0,05011	-2,99353	0,045649	-3,08678
78	0,0523	-2,95076	0,052943	-2,93854	0,050705	-2,98173
79	0,058231	-2,84334	0,064098	-2,74735	0,056982	-2,86503
80	0,065146	-2,73112	0,066895	-2,70463	0,063198	-2,76149
81	0,072053	-2,63035	0,082686	-2,4927	0,070698	-2,64933
82	0,084247	-2,474	0,084815	-2,46728	0,078243	-2,54794

83	0,093344	-2,37146	0,096643	-2,33673	0,087424	-2,43699
84	0,102575	-2,27716	0,103285	-2,27026	0,097547	-2,32742
85	0,115623	-2,15742	0,116722	-2,14796	0,109032	-2,21612
86	0,129233	-2,04614	0,128957	-2,04828	0,121649	-2,10662
87	0,143547	-1,9411	0,147185	-1,91606	0,135173	-2,0012
88	0,159574	-1,83525	0,157547	-1,84803	0,150767	-1,89202
89	0,174328	-1,74682	0,182393	-1,70159	0,166034	-1,79556
90	0,194478	-1,63744	0,194839	-1,63558	0,184587	-1,68963
91	0,220244	-1,51302	0,243137	-1,41413	0,203998	-1,58964
92	0,234653	-1,44965	0,246064	-1,40216	0,227229	-1,4818
93	0,268278	-1,31573	0,274921	-1,29127	0,249096	-1,38992
94	0,277045	-1,28358	0,304884	-1,18782	0,273847	-1,29519
95	0,323917	-1,12727	0,336935	-1,08786	0,302268	-1,19644
96	0,349977	-1,04989	0,331319	-1,10467	0,326274	-1,12002
97	0,361471	-1,01757	0,371167	-0,9911	0,351432	-1,04574
98	0,398023	-0,92125	0,339555	-1,08012	0,367485	-1,00107
99	0,44648	-0,80636	0,387881	-0,94706	0,392321	-0,93567

## B.2. Análisis en la población masculina para el año 2011

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,00323	-5,73519	0,002295	-6,07701	0,002162	-6,13651
1	0,000279	-8,18308	0,00016	-8,73906	0,000251	-8,28835
2	0,000179	-8,62937	0,000116	-9,05939	0,000222	-8,41506
3	0,00014	-8,87367	0,000106	-9,15435	0,00015	-8,80274
4	8,94E-05	-9,32225	9,23E-05	-9,29071	0,000123	-9,00691
5	8,66E-05	-9,35391	8,86E-05	-9,3318	0,000109	-9,12249
6	0,000124	-8,9959	9,5E-05	-9,26152	0,000105	-9,1662
7	8,59E-05	-9,36232	8,88E-05	-9,32863	0,000101	-9,19962
8	7,52E-05	-9,49597	9,1E-05	-9,30429	9,86E-05	-9,22398
9	0,000112	-9,09688	8,15E-05	-9,41502	0,000111	-9,10608
10	0,000112	-9,09588	8,57E-05	-9,36482	0,000103	-9,18259
11	6,53E-05	-9,63596	8,89E-05	-9,32827	0,000112	-9,09903
12	7,64E-05	-9,47982	0,000102	-9,19413	0,00012	-9,02788
13	0,000109	-9,1197	0,00013	-8,94649	0,00013	-8,94672
14	0,000214	-8,4515	0,000165	-8,71212	0,000144	-8,84669
15	0,000143	-8,85533	0,000239	-8,34109	0,000188	-8,58071
16	0,000174	-8,65776	0,000378	-7,88162	0,000249	-8,29925
17	0,000247	-8,30411	0,000492	-7,61642	0,000284	-8,16495
18	0,000363	-7,92243	0,000563	-7,48167	0,000354	-7,94602
19	0,000383	-7,8681	0,000638	-7,35786	0,000388	-7,85475
20	0,000435	-7,7393	0,000678	-7,29585	0,00039	-7,84886

21	0,000509	-7,58242	0,000702	-7,26163	0,000414	-7,7894
22	0,000389	-7,85107	0,000709	-7,251	0,000418	-7,78091
23	0,000379	-7,87903	0,000712	-7,24794	0,00042	-7,77459
24	0,000424	-7,76515	0,000683	-7,28959	0,000438	-7,73416
25	0,000429	-7,75447	0,000717	-7,23978	0,000428	-7,75576
26	0,000391	-7,84776	0,000715	-7,24322	0,000441	-7,72656
27	0,00046	-7,68533	0,000758	-7,18462	0,000448	-7,71153
28	0,000466	-7,67127	0,000766	-7,17381	0,00045	-7,70668
29	0,000489	-7,62384	0,000877	-7,03912	0,000474	-7,65405
30	0,000549	-7,50709	0,000891	-7,02322	0,000492	-7,61722
31	0,00054	-7,52389	0,000998	-6,90988	0,000534	-7,5351
32	0,000506	-7,58921	0,000987	-6,92108	0,00056	-7,4883
33	0,000639	-7,35539	0,001055	-6,85432	0,000581	-7,4506
34	0,000597	-7,42379	0,001108	-6,80499	0,000656	-7,32878
35	0,00071	-7,25022	0,001163	-6,75712	0,000688	-7,28119
36	0,000722	-7,2331	0,001221	-6,70791	0,000765	-7,17574
37	0,000862	-7,05638	0,001308	-6,63946	0,000825	-7,10034
38	0,000852	-7,06785	0,001372	-6,59118	0,000941	-6,96877
39	0,001016	-6,89168	0,001573	-6,45477	0,001024	-6,88416
40	0,001178	-6,7436	0,001556	-6,46561	0,001175	-6,74688
41	0,001267	-6,67098	0,001808	-6,31545	0,001297	-6,6478
42	0,001424	-6,55437	0,001782	-6,33023	0,001476	-6,51829
43	0,001687	-6,38455	0,002081	-6,17494	0,001667	-6,39657
44	0,001775	-6,33383	0,002233	-6,10443	0,001871	-6,2814
45	0,002028	-6,20051	0,002423	-6,02287	0,002093	-6,16908
46	0,002285	-6,08146	0,002697	-5,91578	0,00234	-6,05746
47	0,002667	-5,92681	0,002981	-5,81542	0,002658	-5,93022
48	0,002879	-5,85022	0,003165	-5,75547	0,002968	-5,81979
49	0,00317	-5,75409	0,003619	-5,62164	0,003268	-5,72372
50	0,003724	-5,59296	0,003723	-5,59335	0,003667	-5,60825
51	0,003854	-5,55874	0,004382	-5,43033	0,004032	-5,51355
52	0,004606	-5,38037	0,00445	-5,41492	0,004489	-5,40608
53	0,00469	-5,36224	0,005003	-5,29771	0,004849	-5,32898
54	0,005431	-5,21556	0,005436	-5,21469	0,005373	-5,22635
55	0,005569	-5,19057	0,005883	-5,13568	0,005793	-5,15119
56	0,006565	-5,026	0,006423	-5,04789	0,006279	-5,07052
57	0,006898	-4,97654	0,00713	-4,94342	0,006879	-4,97932
58	0,007441	-4,90076	0,007484	-4,89502	0,007439	-4,90107
59	0,008001	-4,82822	0,008609	-4,75495	0,007979	-4,8309
60	0,008759	-4,73771	0,008947	-4,71646	0,008676	-4,74723
61	0,009275	-4,68046	0,010537	-4,55291	0,009289	-4,67897
62	0,010284	-4,57721	0,010521	-4,55441	0,010032	-4,602
63	0,011213	-4,49064	0,011591	-4,45753	0,010893	-4,51961
64	0,011923	-4,4293	0,012926	-4,34854	0,011794	-4,44016

65	0,013067	-4,33769	0,01387	-4,27802	0,012758	-4,3616
66	0,013851	-4,27941	0,015684	-4,15513	0,013811	-4,28228
67	0,015337	-4,1775	0,017064	-4,07076	0,014989	-4,20043
68	0,016433	-4,10848	0,018503	-3,98983	0,016314	-4,11572
69	0,017938	-4,02082	0,021567	-3,8366	0,017777	-4,02985
70	0,01984	-3,92005	0,021745	-3,82836	0,019675	-3,92839
71	0,02156	-3,83691	0,026451	-3,63246	0,021448	-3,84212
72	0,024484	-3,70974	0,026627	-3,62584	0,02387	-3,73514
73	0,027423	-3,59636	0,03024	-3,49857	0,026373	-3,63541
74	0,029933	-3,5088	0,03332	-3,40159	0,029219	-3,53295
75	0,031365	-3,46206	0,036917	-3,29908	0,032436	-3,42848
76	0,037008	-3,29662	0,041625	-3,17906	0,035927	-3,32626
77	0,040397	-3,20899	0,046955	-3,05856	0,040248	-3,21268
78	0,045357	-3,09318	0,049006	-3,01582	0,044961	-3,10196
79	0,050458	-2,98662	0,060716	-2,80155	0,050576	-2,98427
80	0,057583	-2,85452	0,062448	-2,77342	0,05665	-2,87086
81	0,066003	-2,71805	0,079557	-2,53128	0,063524	-2,75634
82	0,073085	-2,61613	0,079836	-2,52777	0,071026	-2,64471
83	0,081807	-2,50339	0,091946	-2,38655	0,079436	-2,5328
84	0,08998	-2,40816	0,097182	-2,33116	0,088438	-2,42545
85	0,102781	-2,27516	0,110915	-2,19899	0,098623	-2,31645
86	0,114285	-2,16906	0,12247	-2,09989	0,110806	-2,19998
87	0,128795	-2,04953	0,140997	-1,95901	0,124421	-2,08408
88	0,144623	-1,93363	0,150016	-1,89702	0,140203	-1,96467
89	0,158474	-1,84216	0,176383	-1,7351	0,155916	-1,85844
90	0,182705	-1,69988	0,187146	-1,67586	0,174939	-1,74332
91	0,201416	-1,60238	0,243139	-1,41412	0,194324	-1,63823
92	0,220813	-1,51044	0,240482	-1,42511	0,216155	-1,53176
93	0,249012	-1,39025	0,271626	-1,30333	0,240809	-1,42375
94	0,264756	-1,32895	0,301851	-1,19782	0,264206	-1,33103
95	0,296165	-1,21684	0,337067	-1,08747	0,289766	-1,23868
96	0,328296	-1,11384	0,324943	-1,12411	0,311855	-1,16522
97	0,346252	-1,06059	0,368897	-0,99724	0,343448	-1,06872
98	0,372639	-0,98714	0,326377	-1,1197	0,362227	-1,01548
99	0,435519	-0,83122	0,375507	-0,97948	0,388071	-0,94657

### B.3. Análisis en la población masculina para el año 2016

Edad	$m_{x,t}$	$\ln(m_{x,t})$	LC $m_{x,t}$	LC $\ln(m_{x,t})$	LSTM $m_{x,t}$	LSTM $\ln(m_{x,t})$
0	0,003002	-5,80858	0,001617	-6,42726	0,001839	-6,29869
1	0,000291	-8,14226	0,000111	-9,10714	0,000229	-8,38363
2	0,000113	-9,08759	8,23E-05	-9,40512	0,000173	-8,65941

3	7,05E-05	-9,55995	7,75E-05	-9,46508	0,000113	-9,08719
4	0,0001	-9,21009	6,81E-05	-9,59385	0,000107	-9,14587
5	9,71E-05	-9,2396	6,67E-05	-9,61478	0,000102	-9,18783
6	7,66E-05	-9,47657	7,31E-05	-9,52321	9,22E-05	-9,29128
7	8,45E-05	-9,37914	6,88E-05	-9,58462	9,23E-05	-9,29087
8	6,48E-05	-9,64405	7,17E-05	-9,54281	9,24E-05	-9,28934
9	8,2E-05	-9,40878	6,38E-05	-9,66024	8,73E-05	-9,34609
10	7,89E-05	-9,44703	6,77E-05	-9,6	8,79E-05	-9,33952
11	8,81E-05	-9,33701	7,11E-05	-9,55114	8,79E-05	-9,33978
12	8,19E-05	-9,40969	8,22E-05	-9,40677	0,0001	-9,20957
13	8,34E-05	-9,3916	0,000108	-9,1341	0,000103	-9,17996
14	9,02E-05	-9,31313	0,000137	-8,89339	0,000111	-9,10381
15	0,000133	-8,92508	0,000206	-8,48532	0,00014	-8,87736
16	0,000216	-8,43796	0,00034	-7,98612	0,00016	-8,73746
17	0,000156	-8,76605	0,000449	-7,70857	0,000198	-8,52565
18	0,000239	-8,34106	0,000511	-7,57917	0,000238	-8,34402
19	0,000344	-7,97614	0,000582	-7,44889	0,000273	-8,20434
20	0,000315	-8,06165	0,00062	-7,38629	0,000294	-8,13079
21	0,000255	-8,27333	0,000644	-7,34734	0,000302	-8,10626
22	0,000356	-7,94027	0,00065	-7,3384	0,00031	-8,07857
23	0,000383	-7,86648	0,00065	-7,33911	0,00032	-8,04673
24	0,000309	-8,08129	0,000614	-7,39631	0,00034	-7,98768
25	0,00029	-8,14525	0,000647	-7,34337	0,000359	-7,93281
26	0,000433	-7,74394	0,000643	-7,34879	0,000362	-7,92313
27	0,000432	-7,74707	0,000684	-7,28771	0,000386	-7,85863
28	0,000386	-7,85981	0,000689	-7,28	0,000391	-7,84631
29	0,000396	-7,83355	0,000806	-7,12376	0,000407	-7,80662
30	0,000552	-7,50263	0,000811	-7,11702	0,00042	-7,77451
31	0,000503	-7,59591	0,000928	-6,98242	0,00045	-7,70627
32	0,000471	-7,66131	0,000902	-7,01114	0,000463	-7,67839
33	0,000573	-7,46519	0,000969	-6,93975	0,000517	-7,56777
34	0,000576	-7,46003	0,001023	-6,88532	0,000546	-7,5125
35	0,000555	-7,49655	0,001071	-6,83894	0,000586	-7,44242
36	0,000671	-7,30685	0,001124	-6,79081	0,000628	-7,37277
37	0,000628	-7,37251	0,001207	-6,71943	0,000676	-7,29996
38	0,000669	-7,31014	0,001259	-6,67754	0,00073	-7,22255
39	0,000786	-7,14805	0,001464	-6,52687	0,000806	-7,12289
40	0,000943	-6,96687	0,001419	-6,55751	0,000862	-7,0561
41	0,001028	-6,87993	0,001694	-6,38044	0,000967	-6,94176
42	0,001005	-6,90315	0,001619	-6,42579	0,001095	-6,81709
43	0,001237	-6,69489	0,001933	-6,24889	0,001242	-6,6914
44	0,001396	-6,57396	0,002065	-6,1826	0,001388	-6,5799
45	0,001603	-6,43582	0,002233	-6,10438	0,001605	-6,4344
46	0,001675	-6,39207	0,002504	-5,98976	0,001816	-6,31091
47	0,002042	-6,19398	0,002771	-5,88871	0,002055	-6,18751

48	0,002505	-5,98956	0,002916	-5,83743	0,002349	-6,05391
49	0,00264	-5,93691	0,003377	-5,69085	0,002727	-5,9046
50	0,003061	-5,78903	0,003412	-5,68049	0,003074	-5,78483
51	0,003448	-5,66982	0,004113	-5,49364	0,00344	-5,6723
52	0,003936	-5,53764	0,004083	-5,50088	0,003883	-5,55113
53	0,004295	-5,45034	0,004622	-5,37688	0,004308	-5,44722
54	0,00476	-5,3474	0,005017	-5,29489	0,004692	-5,36189
55	0,005024	-5,29353	0,005432	-5,21543	0,005205	-5,25809
56	0,005714	-5,16481	0,005917	-5,12996	0,005688	-5,16943
57	0,006245	-5,07603	0,006601	-5,02049	0,006326	-5,06303
58	0,006773	-4,99476	0,006852	-4,98319	0,006868	-4,98082
59	0,007503	-4,89249	0,007965	-4,83266	0,007535	-4,88815
60	0,008283	-4,79361	0,008162	-4,80821	0,008241	-4,79862
61	0,009059	-4,70396	0,00985	-4,62033	0,008833	-4,72927
62	0,009396	-4,66747	0,009591	-4,64695	0,009521	-4,65421
63	0,010777	-4,53039	0,010584	-4,54842	0,010234	-4,58207
64	0,011503	-4,46517	0,011827	-4,43738	0,011104	-4,50623
65	0,012422	-4,38826	0,012626	-4,37197	0,01186	-4,4346
66	0,013595	-4,29806	0,014423	-4,23895	0,01295	-4,34667
67	0,0139	-4,27585	0,015629	-4,15862	0,014015	-4,26764
68	0,016237	-4,12047	0,016915	-4,07953	0,015416	-4,17234
69	0,016972	-4,07622	0,020015	-3,91126	0,016578	-4,09969
70	0,018102	-4,01175	0,01971	-3,92661	0,018214	-4,00554
71	0,020508	-3,88695	0,024665	-3,70237	0,019703	-3,92701
72	0,022206	-3,80737	0,024126	-3,72446	0,02165	-3,83274
73	0,024639	-3,70341	0,027642	-3,58842	0,023587	-3,74708
74	0,027092	-3,60852	0,030405	-3,49316	0,02612	-3,64504
75	0,030153	-3,50147	0,033744	-3,38896	0,029319	-3,5295
76	0,033482	-3,39674	0,038204	-3,2648	0,032443	-3,42827
77	0,037041	-3,29574	0,043289	-3,13985	0,036557	-3,30889
78	0,040881	-3,19709	0,044494	-3,11241	0,040766	-3,1999
79	0,046873	-3,0603	0,056738	-2,86931	0,045598	-3,0879
80	0,05108	-2,97437	0,057303	-2,8594	0,050664	-2,98255
81	0,057457	-2,85672	0,075813	-2,57949	0,05672	-2,86963
82	0,065713	-2,72245	0,074022	-2,60339	0,063458	-2,75738
83	0,07208	-2,62998	0,086395	-2,44883	0,071477	-2,63838
84	0,083076	-2,488	0,090059	-2,40729	0,080061	-2,52497
85	0,095125	-2,35257	0,10406	-2,26279	0,090151	-2,40627
86	0,106316	-2,24134	0,114819	-2,1644	0,101225	-2,29041
87	0,118757	-2,13067	0,133627	-2,0127	0,113625	-2,17485
88	0,133713	-2,01206	0,141106	-1,95824	0,127643	-2,05852
89	0,151489	-1,88724	0,169147	-1,77699	0,142747	-1,94668
90	0,169386	-1,77557	0,177956	-1,72622	0,161228	-1,82493
91	0,187254	-1,67529	0,243142	-1,41411	0,180335	-1,71294
92	0,208386	-1,56836	0,233681	-1,4538	0,201995	-1,59951



93	0,228409	-1,47662	0,267561	-1,31841	0,226302	-1,48588
94	0,263005	-1,33558	0,298101	-1,21032	0,247698	-1,39555
95	0,269004	-1,31303	0,337231	-1,08699	0,272963	-1,29842
96	0,29972	-1,20491	0,317145	-1,1484	0,299264	-1,20643
97	0,330435	-1,10735	0,366078	-1,00491	0,329283	-1,11084
98	0,36428	-1,00983	0,310622	-1,16918	0,348954	-1,05281
99	0,37944	-0,96906	0,360592	-1,02001	0,374399	-0,98243