



5.3.3 Construcción, calibración y validación del modelo de regresión.

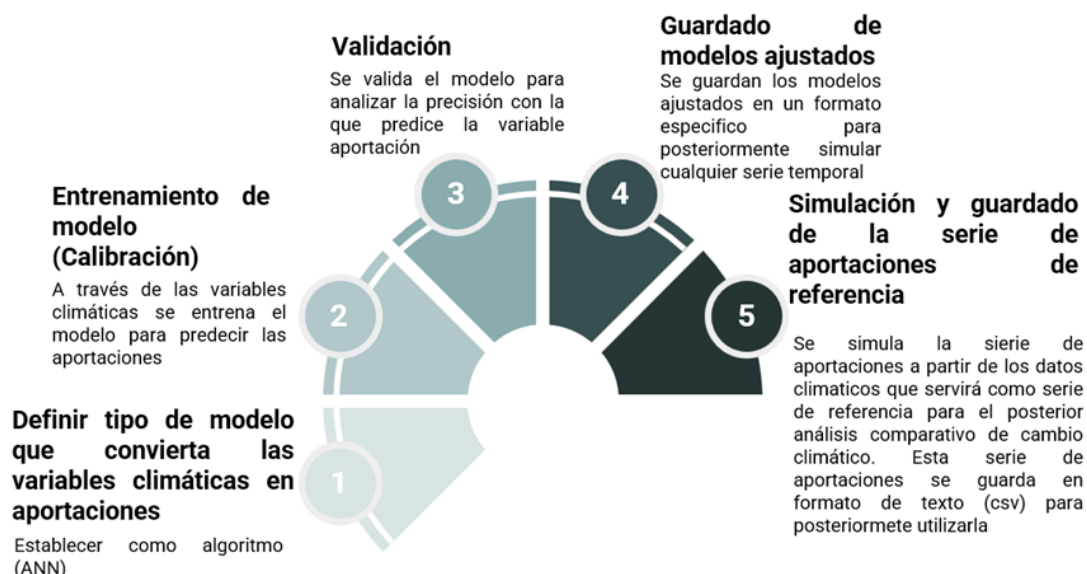
En la metodología que se expone en esta guía se propone la utilización de algoritmos de regresión basados en redes neuronales (ANN) que sustituyan a los modelos hidrológicos. Esto es posible debido a la resolución temporal mensual que se establece en esta metodología. El uso de algoritmos de regresión facilita el proceso iterativo de calibración, ya que dicho proceso es realizado de forma automática y precisa. La elección del algoritmo de ANN frente a los que aparecen en fuentes bibliográficas se debe principalmente al tiempo computacional necesario para ajustar los modelos y a la alta predictibilidad que se consigue en las aportaciones.

Una vez definido el algoritmo de regresión que se utiliza en la metodología, es necesario especificar cuáles son las variables predictoras (variables a partir de las cuales se predice) y cuál es el predictando (variable que se predice) en la configuración de cualquiera de los algoritmos.

En este tipo de estudios las principales variables predictoras son las variables climáticas precipitación, temperatura máxima y temperatura mínima, a resolución mensual. Estas variables permiten configurar los modelos de redes neuronales para obtener como resultado las aportaciones de una cuenca vertiente en un punto dado. En estos modelos es necesario realizar un entrenamiento (calibración) a partir de series de aportaciones previamente obtenidas de fuentes de datos u observaciones.

A continuación, en la Figura 24 se muestra el proceso a seguir para configurar, calibrar y validar los modelos de redes neuronales que permitan reproducir las aportaciones en un periodo de tiempo a partir de las variables climáticas.

Figura 24: Proceso de construcción, calibración y validación de modelos de regresión.



Fuente: Elaboración propia.

En la Tabla 11 se muestran los datos de partida que son necesarios para realizar el proceso descrito anteriormente y los resultados que se obtienen.



Tabla 11. Inputs de partida y resultados del proceso de configuración, calibración y validación de modelos de regresión.

Inputs de partida		Resultados
Tipo de algoritmo	Series temporales de las variables climáticas	Ficheros de configuración de los modelos para posteriormente ser utilizados
Serie temporal de aportaciones		

1. Definir el tipo de modelo que convierta las variables climáticas en series de aportaciones

El primer paso a seguir es la elección del tipo de modelo que convierta las variables climáticas en aportaciones como puede ser un modelo hidrológico o como en esta metodología, modelos de regresión. Como se ha comentado anteriormente en esta metodología se propone el uso de ANN (redes neuronales artificiales).

2. Entrenamiento del modelo

Una vez que se tiene definido el tipo de modelo que se desea utilizar, se requiere definir las variables climáticas necesarias; en la metodología descrita en esta guía se utiliza la precipitación, la temperatura máxima mensual media (la media de las temperaturas máximas del mes) y la temperatura mínima mensual media (la media de las temperaturas mínimas del mes).

Una vez definidas las variables climáticas que se utilizan, es preciso estructurar tanto los predictores como el predictando en forma de matriz para que el modelo de redes neuronales pueda ejecutarse correctamente. La configuración matricial consiste en unir todos los datos climáticos de los puntos de la cuenca extraídos en el apartado 5.3.2.

En la expresión (5) se muestra la configuración matricial necesaria en la que en las filas se encuentran los valores de precipitación y temperatura de cada mes del año (t) y en las columnas los puntos distribuidos en la cuenca (n). En la columna final se encuentra la serie temporal de las aportaciones del punto de estudio.

$$\begin{array}{ccccccc}
 P_{1,1} & T_{max_{1,1}} & T_{min_{1,1}} & \dots & P_{1,n} & T_{max_{1,n}} & T_{min_{1,n}} \longrightarrow Aport_1 \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 P_{t,1} & T_{max_{t,1}} & T_{min_{t,1}} & \dots & P_{t,n} & T_{max_{t,n}} & T_{min_{t,n}} \longrightarrow Aport_t
 \end{array} \quad (5)$$

En la utilización de inteligencia artificial uno de los problemas que pueden existir es el sobreajuste u *overfitting*, en pocas palabras, significa tener en cuenta demasiada información de sus datos y / o conocimientos previos, y utilizarla en un modelo.

En el caso de la utilización de puntos distribuidos en las cuencas donde se introduce la información climática que permitan reproducir las aportaciones, dan lugar a sobreajustes cuando el área de la cuenca tiene un tamaño considerable. Por esta razón es necesario redimensionar la matriz que se muestra en la expresión (5) a través de componentes principales (PCA). *Principal Component Analysis* (PCA) es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez que conserva su información. En esta metodología se utiliza el método PCA que explique el 95% de la varianza del conjunto de los datos, por lo que la matriz de la expresión (5) quedaría reducida a un número de columnas inferior que representan el conjunto de los datos considerados inicialmente.

Configurados los datos de la forma descrita anteriormente y reducida la dimensión, se proceden a entrenar y validar el modelo de redes neuronales.



La forma de entrenar o calibrar un modelo de redes neuronales es a través de la búsqueda de la mejor configuración de parámetros para minimizar o maximizar la métrica deseada. En las herramientas desarrolladas en esta guía se trata de maximizar el coeficiente de determinación (R^2), maximizar el coeficiente Nash-Sutcliffe (NSE) y minimizar el porcentaje de sesgo (PBIAS), en los siguientes apartados se explica cada una de estas métricas.

El proceso de optimización consta de dos fases, una fase propiamente de *entrenamiento* a partir del 80% de los datos de aportaciones desde el año 1950 hasta 2015, y una fase de *prueba* que utiliza el 20% de los datos para verificar el ajuste realizado. Con este proceso se consiguen generar modelos que reproduzcan de forma fidedigna las series de aportaciones en el punto de estudio considerado.

3. Validación de los modelos

Al ser una variable hidrológica lo que se obtiene como resultado de los modelos, es necesario realizar un análisis para validar que el modelo reproduce el comportamiento hidrológico de forma adecuada. Para ello se utilizan dos criterios basados en métricas como son:

- Coeficiente de eficiencia de Nash-Sutcliffe a nivel mensual (NSE) (D. N. Moriasi et al., 2007)
- Porcentaje de sesgo (PBIAS)

Una vez que el modelo está configurado se simula un periodo de tiempo entre 5 y 10 años para poder ser comparada la serie resultante con la serie adoptada como real.

El coeficiente de eficiencia NSE es uno de los más usados en hidrología para la comprobación de que los ajustes simulados son aceptables. El rango de variación del citado coeficiente de eficiencia se sitúa entre $-\infty$ y 1, correspondiendo este último valor a un ajuste perfecto. Con respecto a la medida de la desviación del volumen, expresa cuantitativamente la relación entre el volumen del hidrograma observado y el simulado. El ajuste perfecto corresponde a un valor de PBIAS = 0, por lo que se debe obtener el valor más bajo del mismo. Esta media de bondad de ajuste es adecuada para análisis mensuales y anuales de aportaciones.

Valores positivos del coeficiente PBIAS, indican infraestimación de los valores de aportaciones simulados por el modelo. Por el contrario, valores negativos del mismo indican sobreestimación de las aportaciones simuladas por el modelo con respecto a los observados o tomados como referencia (D. N. Moriasi et al., 2007).

Los mencionados coeficientes se determinan a partir de las siguientes expresiones:

$$NSE = 1.0 - \frac{\sum_{i=1}^n (Aport_{0i} - Aport_{si})^2}{\sum_{i=1}^n (Aport_{0i} - Aport_{0m})^2} \quad (6)$$

$$PBIAS = \frac{\sum_{i=1}^n (Aport_{0i} - Aport_{si})}{\sum_{i=1}^n (Aport_{0i})} * 100 \quad (7)$$

NSE: Coeficiente de eficiencia de Nash-Sutcliffe.

PBIAS: Desviación del volumen.

N: Número de parejas de caudales consideradas en el proceso de calibración.

$Aport_{0i}$: Aportaciones observadas en el instante i.

$Aport_{si}$: Aportaciones simuladas en el instante i.



$Aport_{0m}$: Valor medio de los caudales observados a lo largo del período de calibración.

En función de la combinación del valor de dichos coeficientes (NSE y PBIAS) conjuntamente, se pueden distinguir distintos niveles de calidad del ajuste, según se indica en la tabla siguiente:

Tabla 12. Niveles de calidad del ajuste del nivel diario, en función del valor del coeficiente NSE y PBIAS.

Calidad del ajuste	NSE	PBIAS (%)
Excelente	$0,75 < NSE \leq 1,00$	$PBIAS < \pm 10$
Bueno	$0,65 < NSE \leq 0,75$	$\pm 10 \leq PBIAS < \pm 15$
Aceptable	$0,50 < NSE \leq 0,65$	$\pm 15 \leq PBIAS < \pm 25$
No aceptable	$NSE \leq 0,50$	$PBIAS \geq \pm 25$

Nota: conforme a Moriasi et al., 2007.

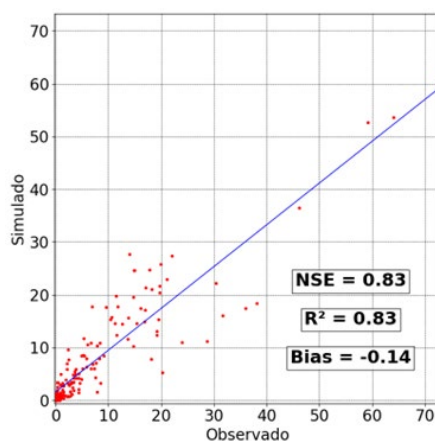
Otros de los coeficientes que se analizan es el coeficiente de determinación (R^2), medida estadística que define lo cerca están los datos de la línea de regresión ajustada y que refleja la bondad del ajuste de un modelo a la variable que pretender explicar. Es importante saber que el resultado del coeficiente de determinación oscila entre 0 y 1. Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar. De forma inversa, cuanto más cerca de cero, menos ajustado estará el modelo y, por tanto, menos fiable será.

$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} \quad (8)$$

Donde σ_{XY}^2 es la covarianza entre las aportaciones simuladas y las aportaciones observadas, σ_X^2 la varianza de las aportaciones observadas y σ_Y^2 la varianza de las aportaciones simuladas

Como resultado de validación del modelo calibrado, la aplicación de la herramienta de esta guía proporciona los coeficientes NSE, R^2 y PBIAS, utilizando para su cálculo el 20% de los datos restantes de la serie de aportaciones utilizada para calibrar el modelo, en este caso la de SIMPA, y las aportaciones resultantes de la red neuronal. Además, se obtiene de forma gráfica el grado de ajuste entre los resultados observados y simulados como se muestra en la Figura 25.

Figura 25: Validación de resultados obtenidos en la cuenca vertiente a la presa de Santillana (Madrid).

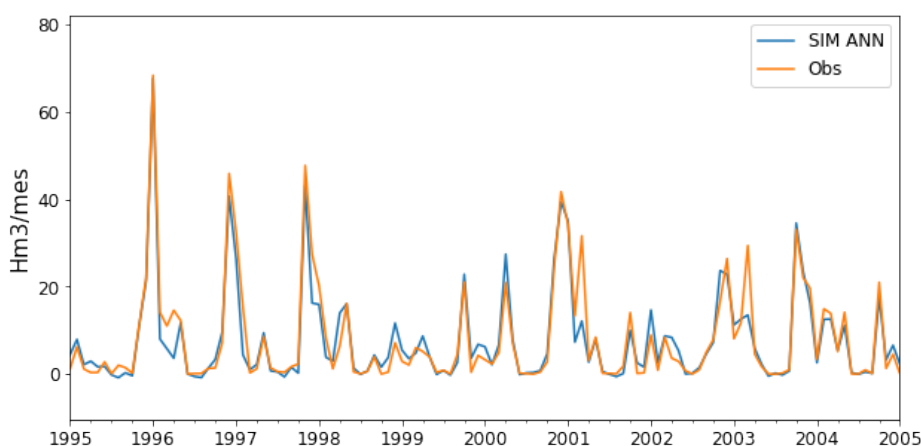


Fuente: Elaboración propia.



Además, a través de las herramientas proporcionadas con esta guía es posible comparar la serie temporal de forma gráfica, lo que permite identificar si el modelo reproduce la estacionalidad correctamente. En la Figura 26 se muestra un ejemplo de la comparativa entre la serie temporal de aportaciones simulada y la serie real. Como puede observarse los resultados obtenidos en este ejemplo tomando 5 años de análisis son de excelente calidad ya que se ajustan de forma adecuada a la serie de referencia de SIMPA.

Figura 26: Validación de serie temporal obtenida en la cuenca vertiente a la presa de Santillana (Madrid).



Fuente: Elaboración propia.

4. Guardado de los modelos ajustados

Tras calibrar y validar los modelos, éstos se guardan en un formato determinado que permita posteriormente cargarlos para simular cualquier escenario climático. El tipo de fichero utilizado para guardar los parámetros ajustados de modelos de redes neuronales es el formato ".sav", una extensión genérica que se utiliza para guardar los parámetros de ajuste de los modelos de regresión. Este tipo de ficheros permite cargar los modelos ajustados de forma automática e introducir en formato matricial las series temporales climáticas de los puntos distribuidos en la cuenca.

5. Simulación de la serie de aportaciones de referencia

Una vez configurado el modelo de redes neuronales es necesario realizar la simulación de la serie de aportaciones que servirá como referencia en el posterior análisis de cambio climático. A esta serie se le denomina serie histórica o serie de control. Cualquier análisis comparativo que se realice posteriormente con las series de cambio climático, para analizar los cambios que se producen con respecto al período histórico debe considerarse esta serie acotada al número de años definido en los diferentes periodos futuros, que en esta metodología se propone que sean 30 años (1976-2005).

5.3.4 Actualización de series climáticas en la cuenca de estudio con la información de escenarios de cambio climático

Para poder evaluar cómo va a influir el cambio climático en la temperatura y la precipitación, es necesario utilizar el conjunto de las diferentes predicciones climáticas que permitan reproducir, a través de los modelos que transforman datos climáticos en aportaciones, cada uno de los modelos y escenarios de cambio climático.

En los estudios de cambio climático se debe tener en cuenta el amplio abanico de predicciones que proporcionan los modelos de cambio climático para cada uno de los escenarios de emisiones como se puede observar en la Figura 27.