

DAM:

Data Analysis for Monitoring



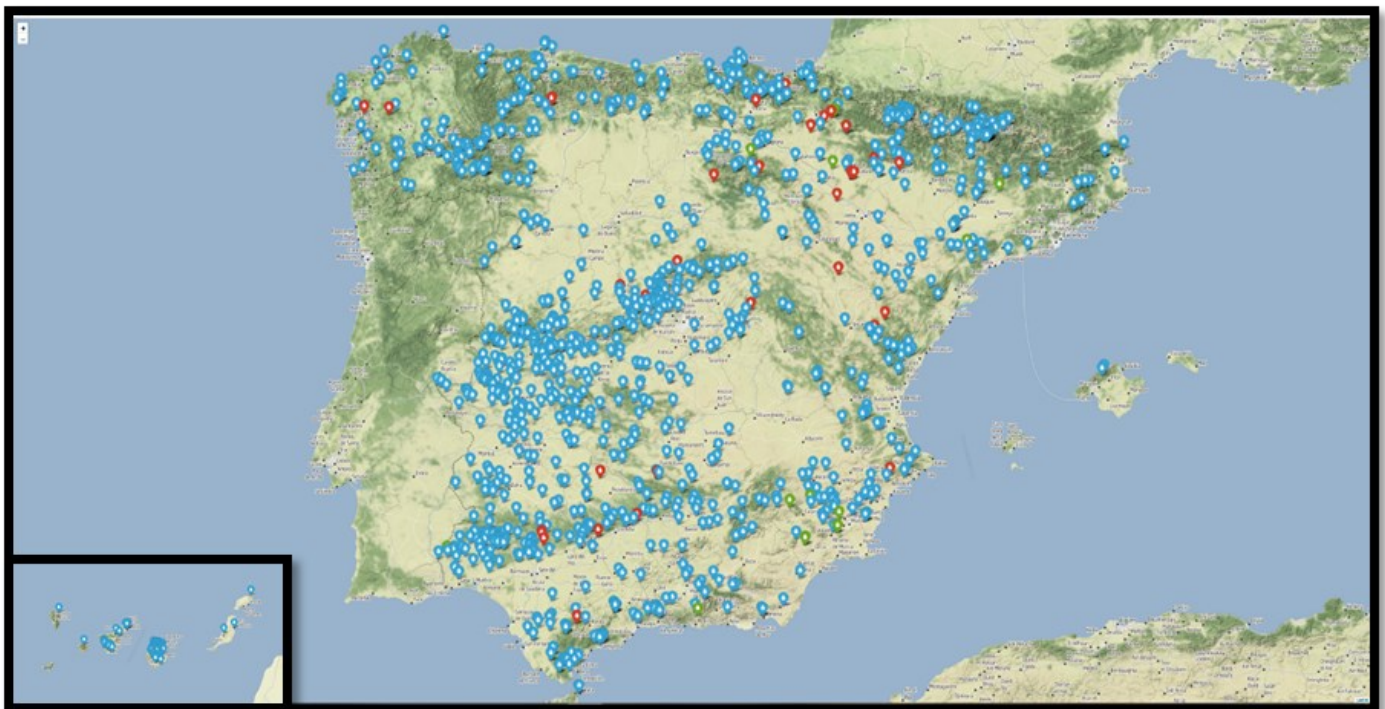
Laura Alonso Cea
Neoland

Data Science Bootcamp abril - junio 2020

Introducción

Según el inventario de la SEPREM, la Sociedad Española de Presas y Embalses, España cuenta en la actualidad con 1226 grandes presas, que en su conjunto suman una capacidad aproximada de 56 000 hm³. De hecho, según los datos de la ICOLD (International Commission on Large Dams), España es del noveno país del mundo con mayor número de estas estructuras. En el momento de la redacción de este proyecto, 38 de las presas del inventario antes señalado se encuentran en fase de construcción, 15 atraviesan la fase de puesta en carga y 1173 están en fase de explotación.

La enorme proliferación de este tipo de estructuras en nuestro país se debe principalmente a un régimen hidrológico extraordinariamente irregular, caracterizado por la alternancia de periodos de sequía prolongados e inundaciones, como manifestaciones extremas de esta gran variabilidad. A este respecto, las presas hidráulicas desempeñan dos funciones esenciales: por un lado, permiten almacenar el agua para poder disponer de ella con fines de consumo humano y regadío durante las épocas de escasez, y, por otro lado, regulan el caudal de los ríos y laminan las avenidas de agua, que constituyen el riesgo natural más importante en España, con 2450 inundaciones en los últimos 500 años. Otra utilidad de las presas radica en su explotación con fines de aprovechamiento hidroeléctrico.



Mapa creado con Folium y datos obtenidos mediante web scraping del sitio de la SEPREM

Tipos de presas

Las presas hidráulicas pueden clasificarse en dos grandes grupos en función de los materiales empleados para su construcción: **presas de fábrica** (normalmente de hormigón) y **presas de materiales sueltos**.

En el primer caso, una pantalla de hormigón levantada sobre la cerrada de una vaguada es el elemento que contiene el agua del embalse. Este tipo de presas pueden adoptar diversas formas: de gravedad, de arco, de contrafuertes, etc.



La Almendra, Salamanca (fuente: www.spancold.org)

Por su parte, el cuerpo de las presas de materiales sueltos se compone de materiales naturales (escollera, gravas, rocas, arenas, limos o arcillas) que no se someten a ningún proceso químico de transformación para su puesta en obra y cuya colocación se realiza mediante procedimientos de compactación. Estos materiales se pueden obtener de la propia excavación del vaso del embalse. Las presas de materiales sueltos pueden dividirse en tres grandes clases: presas homogéneas (el material impermeable conforma la totalidad del cuerpo de la presa), presas de núcleo (compuestas por un núcleo impermeable situado entre dos espaldones de materiales sueltos permeables) y presas de pantalla (compuestas en su mayor parte por materiales permeables, pero con una “pantalla” impermeable apoyada sobre el espaldón de aguas arriba).

El objeto de estudio de este proyecto concreto se centra en el caso de dos presas de materiales sueltos.



Canales, Granada (fuente: www.spancold.org)

Seguridad de las presas

De las 1226 presas actualmente existentes en España, aproximadamente 450 son anteriores a 1960 y más de 100 ya existían en el año 1915. Si bien la seguridad es primordial en cualquier construcción de semejante envergadura, las consecuencias catastróficas que tendría la rotura de una presa, sumadas a la antigüedad de muchas de estas estructuras, la convierten un aspecto absolutamente fundamental en este campo. La necesidad de invertir en el control y conservación de las presas, para poder mantenerlas en unas condiciones de explotación y seguridad acordes a las exigencias modernas, queda, en consecuencia, fuera de toda duda.

Los sistemas de auscultación

Por suerte, hoy en día, los responsables de la gestión de las presas cuentan con modernos sistemas de vigilancia y auscultación que permiten comprender el comportamiento de la presa y sus componentes mediante parámetros físicos, así como identificar y detectar de manera temprana los procesos evolutivos que pueden desembocar en un fallo.

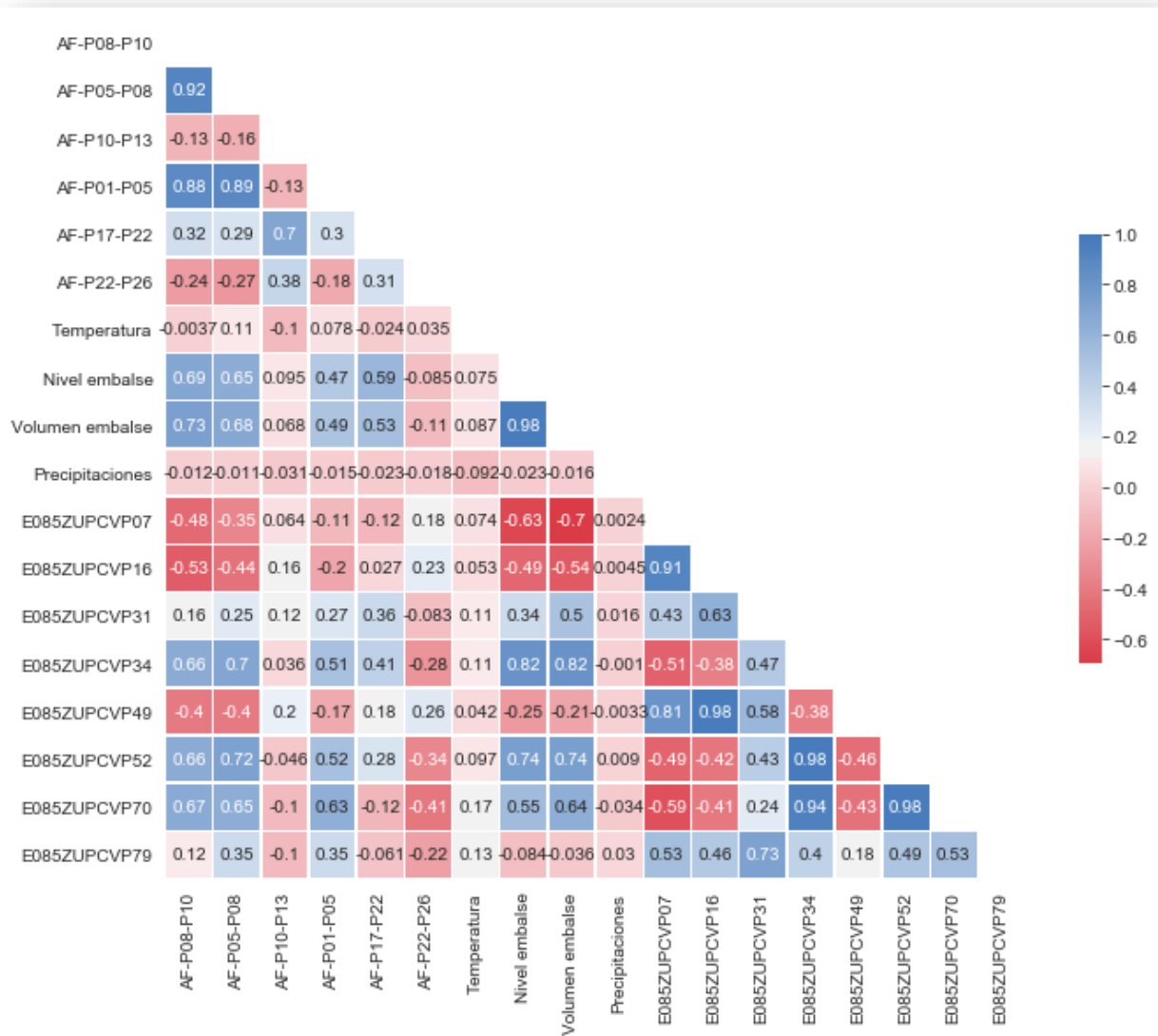
Las variables que miden los dispositivos de los sistemas de auscultación se pueden englobar dentro de dos categorías: variables exteriores y variables respuesta.

Las **variables exteriores**, como su propio nombre indica, son ajenas al comportamiento de la presa y guardan relación, fundamentalmente, con fenómenos meteorológicos. Las más importantes incluyen el nivel del embalse (normalmente aguas arriba, aunque, en algunos casos, también se puede medir el nivel del cauce aguas abajo), la temperatura ambiente y las precipitaciones.

Las **variables respuesta**, por otro lado, reflejan el comportamiento de la presa frente a los esfuerzos a los que ve sometida su estructura. Estas se pueden agrupar, a su vez, en dos clases fundamentales: variables hidráulicas y variables deformacionales. Las primeras, las más importantes en el caso de las presas de materiales sueltos, guardan relación con la acción del agua sobre la presa y la permeabilidad de esta última. Comprenden las **filtraciones** y las **presiones intersticiales**, y se miden mediante **aforadores** y **piezómetros**, respectivamente. Las variables deformacionales, que reflejan desplazamientos de la presa tanto verticales como horizontales, se miden con una gran diversidad de instrumentos o sensores: clavos de nivelación, inclinómetros, células hidráulicas, extensómetros, etc. No obstante, el presente proyecto se centra en el estudio de las variables de naturaleza hidráulica y su relación con las variables exteriores.



En el caso de las presas objeto de esta investigación, la variable exterior que presenta una mayor correlación con las variables hidráulicas es el nivel del embalse, no apreciándose una influencia significativa de la temperatura o las precipitaciones. Así queda reflejado en el mapa de calor que se muestra más abajo, que ilustra la correlación entre las variables exteriores, las presiones intersticiales y los niveles de filtraciones registrados en los aforadores de la presa que, a efectos del presente estudio, se denominará G12. Esto se debe a que se trata de presas construidas con materiales naturales, compactados y con cierta flexibilidad, cuyas características difieren de las del hormigón, por ejemplo, que sí se ve afectado por los cambios en la temperatura como consecuencia del efecto de contracción y dilatación. No obstante, cabe destacar que los efectos de la temperatura sobre las presas de hormigón no tienen consecuencias inmediatas (el cuerpo de la presa tarda un tiempo en calentarse o enfriarse) y, para comprobar dicha correlación, sería preciso cruzar datos con un desfase de aproximadamente un par de meses.



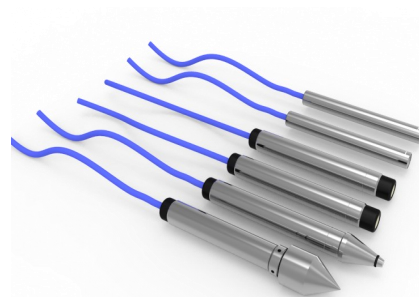
Los instrumentos de medición

Tal y como se ha señalado anteriormente, en los sistemas de auscultación de presas, las variables de naturaleza hidráulica se miden mediante aforadores y piezómetros. A continuación se incluye una breve descripción del principio de funcionamiento de estos instrumentos.



Los aforadores son dispositivos que recogen el caudal de las filtraciones de la presa. Los más habituales y precisos son los denominados aforadores Thompson: estos normalmente cuentan con una placa de acero inoxidable con un rebaje en forma de V, cuya parte inferior provoca una acumulación o remanso de agua que se vierte a través del rebaje hacia aguas abajo del aforador. La medida de la altura del agua sobre el vértice del aforador permite determinar el caudal mediante fórmulas conocidas. Los modelos de lectura manual disponen de una regla graduada de acero inoxidable que, colocada de la forma conveniente, permite ver la altura del agua aguas arriba del citado vértice. Los modelos de lectura automática, además de la placa y la regla, constan de un sensor de ultrasonidos que mide automáticamente la altura del agua y genera una señal analógica que permite su conexión a un sistema automático de adquisición de datos.

Los piezómetros más habituales y precisos son los de cuerda vibrante, que, como su propio nombre indica, disponen en su interior de una cuerda vibrante unida a una membrana inoxidable, permeable y elástica. Cuando se aplica presión hidráulica sobre dicha membrana, se producen deformaciones en la cuerda vibrante, cuya frecuencia varía de manera proporcional a la presión ejercida. Dos bobinas acopladas electromagnéticamente a la cuerda miden esta frecuencia, que se puede leer a través de una unidad portátil o mediante la conexión del piezómetro a un sistema automático de adquisición de datos.



Un caso real: la presa G12

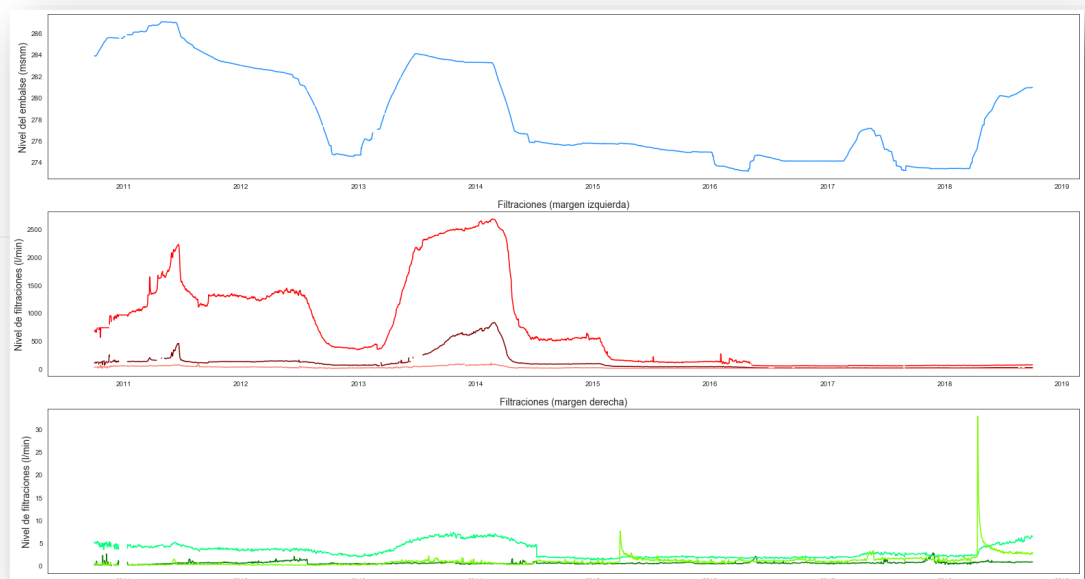
En el presente proyecto, el seudónimo G12 hace referencia a una presa de materiales sueltos, heterogénea, compuesta por un núcleo impermeable de arcilla y espaldones de zahorra. Además, el paramento de aguas arriba está protegido por una capa de escollera y el paramento de aguas abajo dispone de una capa de tierra vegetal hidrosemada que evita la erosión por escorrentía de la lluvia y mimetiza la construcción con su entorno. La altura máxima de esta presa es de 34 m sobre la cimentación.

Once años tras el inicio de los trabajos de construcción de la presa, incluida una campaña de inyecciones de impermeabilización, comenzó la fase de puesta en carga, cuyo primer escalón se alcanzó en cinco meses. Prácticamente un año después, se inició el segundo escalón del plan de puesta en carga de la presa, que debía finalizar con el llenado de esta.

No obstante, a medida que fue subiendo el nivel de embalse, los caudales de filtración aforados en la margen izquierda de la presa se incrementaron notablemente. El nivel del embalse se mantuvo en la misma cota durante aproximadamente un mes y medio y, tras comprobar que los caudales de las filtraciones no disminuían, se decidió proceder a un desembalse parcial.

Cerca de un año y medio después, se decidió permitir que el nivel del embalse se incrementara: los datos del sistema de auscultación revelaron que las filtraciones aforadas en la margen izquierda habían aumentado exponencialmente. Al observar, durante varios meses, que las filtraciones seguían incrementándose aun cuando el nivel del embalse bajaba, se decidió proceder a un nuevo desembalse.

La información registrada por los aforadores de la margen izquierda permitió detectar unos niveles de filtraciones anómalos, derivados, principalmente, de un problema estructural de la presa, motivo por el que se decidió construir una pantalla de impermeabilización continua sobre el estribo izquierdo. La actuación se prolongó durante dos años, aunque en el transcurso de la misma ya se pudo observar cómo las filtraciones habían vuelto a situarse en los niveles existentes al inicio de la fase de puesta en carga para el mismo nivel de embalse. Los gráficos de líneas a continuación reflejan la evolución del nivel del embalse y las filtraciones, tanto de la margen izquierda como derecha, de la presa G12:



DAM: una red neuronal para la predicción de las filtraciones

El ejemplo de la presa G12 demuestra la importancia de los datos de los sistemas de auscultación para detectar anomalías en el comportamiento de este tipo de estructuras y poder llevar a cabo las actuaciones oportunas a tiempo, subsanando de este modo problemas que, en última instancia, podrían provocar la rotura de la presa con desastrosas consecuencias. Teniendo en cuenta todo lo expuesto anteriormente, la posibilidad de predecir los valores de las variables que miden los instrumentos de dichos sistemas reviste un gran interés. Por este motivo, el objetivo último del presente proyecto consiste en la creación de una red neuronal capaz de prever la evolución de las lecturas de unos determinados sensores en función de su historial previo y de los valores de otras variables. La metodología empleada para tal fin se describe a continuación.

1. Datos disponibles:

Para la creación de la red neuronal DAM, se seleccionan los datos de auscultación de una presa de materiales sueltos, denominada en lo sucesivo con el seudónimo E08, que reúne características similares a las de la presa G12, aunque posee mayor antigüedad. La información se divide en tres datasets que contienen las lecturas de las filtraciones, las presiones intersticiales y las variables exteriores respectivamente. Los valores corresponden a las mediciones registradas entre octubre de 2009 y el 18 de junio de 2020.

2. Preparación de los datos:

La información se depura mediante funciones de las librerías **Pandas** y **NumPy** de **Python**. Se eliminan las columnas redundantes, se convierte el formato de los valores de las fechas y se suprimen los datos de los sensores con un elevado porcentaje de valores nulos o que son irrelevantes. La principal dificultad hallada durante este proceso radica en el hecho de que, en los tres datasets de origen, la totalidad de las lecturas se agrupa en una misma columna. Esto obliga a realizar transformaciones del dataframe mediante métodos como *groupby* o *pivot_table* que, en algunos casos, generan una gran cantidad de valores NaN. Cuando el número de sensores no es muy elevado, como en el caso de las mediciones de las variables exteriores, esto se puede subsanar seleccionando únicamente los valores de fecha y medida correspondientes a cada instrumento y fusionándolos en un nuevo dataframe. A lo largo de todo el proceso, se generan gráficos con **Matplotlib** y recuentos que permiten controlar el número de valores NaN.

3. Combinación de los datos de todas las variables:

El objetivo concreto que se persigue con la creación del modelo consiste en predecir, con una antelación de un día, los valores de filtraciones de un determinado aforador (AF-01) a partir de los datos de las variables exteriores y las presiones intersticiales. Por ello, es preciso unir los dataframes de los tres tipos de variables en uno solo. Una vez hecho esto, se procede a eliminar todos los registros con valores NaN para evitar introducir “ruido” en el modelo. El resultado final es un dataframe que contiene 2506 registros y 16 variables (los valores de la temperatura, el nivel del embalse, las precipitaciones, cinco aforadores y ocho piezómetros).

4. Creación de la red:

Mediante el uso de las librerías **Scikit-learn** y **Keras**, se procede a crear un modelo de red neuronal recurrente, por ser este tipo el más adecuado para las series temporales. En primer lugar, se define la función *series_to_supervised*, que permite transformar series temporales en problemas de aprendizaje supervisado (formato necesario para el entrenamiento de la red). A continuación, se normalizan los datos mediante el método *MinMaxScaler* para que adquieran un valor en un rango de entre 0 y 1 y se aplica la función *series_to_supervised* a los datos ya normalizados. Después se divide el dataframe en dos grupos: uno con los datos de entrenamiento, que comprende los primeros 2000 registros, y otro con los datos de prueba, conformado por los 505 registros restantes. Posteriormente se aplica el método *reshape* para que la entrada de la red sea tridimensional. Por último, se genera una red de tipo *Sequential* que cuenta con: una primera capa LSTM de 50 entradas, una segunda capa de 25 neuronas con función de activación ReLU, una tercera capa de 12 neuronas también con función de activación ReLU y una cuarta capa con una salida. Como métrica se configura el error absoluto medio (MAE) y se emplea el optimizador Adam.

```
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM

model = Sequential()
model.add(LSTM(50, input_shape=(train_X.shape[1], train_X.shape[2])))
model.add(Dense(25, activation="relu"))
model.add(Dense(12, activation="relu"))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam')

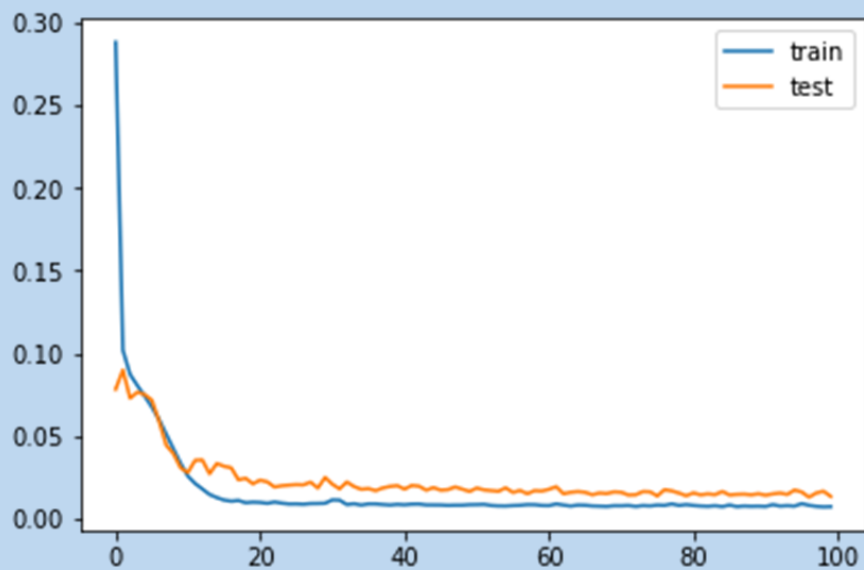
history = model.fit(train_X, train_y, epochs=100, batch_size=100, validation_data=(test_X, test_y), verbose=1)
```

Train on 2000 samples, validate on 505 samples

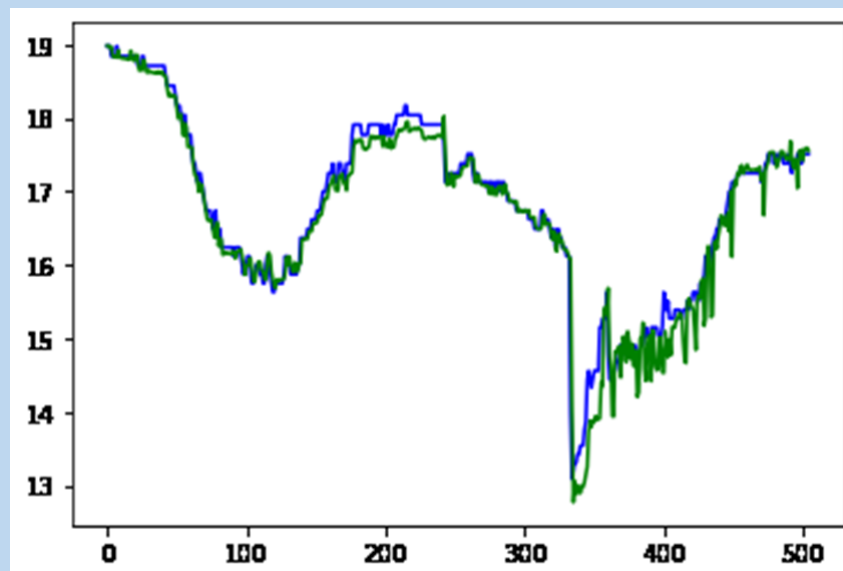
Epoch	Train Samples	Validation Samples	Time per Step	Loss	Val Loss
1/100	2000/2000		1s 479us/step	0.2881	0.0782
2/100	2000/2000		0s 61us/step	0.1019	0.0898
3/100	2000/2000		0s 76us/step	0.0872	0.0728
4/100	2000/2000		0s 60us/step	0.0805	0.0765
5/100	2000/2000		0s 76us/step	0.0742	0.0751
6/100					

5. Resultados:

Tras un entrenamiento de la red con 100 épocas, se genera el siguiente gráfico de disminución del error:



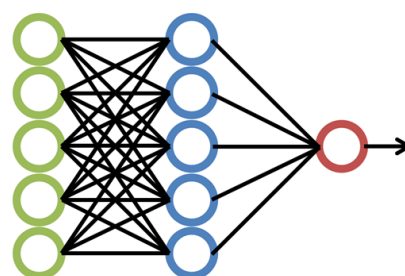
A continuación, se genera la predicción aplicando el modelo al conjunto de datos de prueba. Tras devolver los datos de prueba y de la predicción al formato y rango previos al entrenamiento, se calcula la raíz del error cuadrático medio para validar los resultados y se obtiene un valor de 0.276. Por último, se procede a generar un gráfico donde quedan representados los valores reales (línea azul) y los de la predicción (línea verde).



Conclusiones y futuro

Tal y como se puede apreciar en el gráfico de la página anterior, los valores predichos por el modelo se ajustaron a los valores reales con bastante precisión. Esto demuestra el potencial de la red neuronal DAM para predecir las filtraciones registradas por los sistemas de auscultación de las presas.

Como continuación de este proyecto, sería muy interesante probar la red DAM con otras métricas y plazos de predicción más amplios. Del mismo modo, se podría emplear para predecir los valores de otras variables, como las relativas al control deformacional. El abanico de posibilidades es muy amplio y el uso de esta tecnología, si los resultados presentan un grado de precisión suficiente, podría convertirse en una herramienta de gran utilidad para el control y mantenimiento de las presas en el futuro.



Bibliografía

<https://www.miteco.gob.es/es/>

<https://www.icold-cigb.org/>

<http://www.seprem.es/>

SANTILLÁN, D., FRAILE-ARDANUY, J. & TOLEDO, M.A., Predicción de lecturas de aforos de filtraciones de presas bóveda mediante redes neuronales artificiales. *Tecnología y Ciencias del Agua*. Vol. V, núm. 3, mayo-junio de 2014, pp. 81-96.

TAYFUR, G., SWIATEK, D., WITA, A., and SINGH, V.P., Case Study: Finite Element Method and Artificial Neural Network Models for Flow through Jeziorsko Earthfill Dam in Poland. *Journal of Hydraulic Engineering*. Vol. 131, No. 6, June, 2005, pp. 431-440.