Geophys. J. Int. (2025) **200**, 1–**??**

EIVIA 2025: Machine Learning para Series Temporales con Aplicaciones a Medicina Clínica

Mateo D. Williams[1], Matias Cisnero[2]
[1] *mateodelgadowilliams@gmail.com,* [2] *matiasnicolas.cisnero@estudiantes.unahur.edu.ar*

8 April 2025

SUMMARY
Este informe presenta un estudio sobre la aplicación de redes neuronales recurrentes (RNN) en la predicción de series temporales, específicamente en el contexto de datos meteorológicos. Se utilizan modelos de aprendizaje automático para predecir la temperatura en un intervalo de 24 horas a partir de diversas magnitudes atmosféricas.

Key words: RNN's, Time Series, Machine learning

1 INTRODUCTION

En el campo del aprendizaje automático y la inteligencia artificial, las redes neuronales recurrentes (RNN, por sus siglas en inglés) han emergido como una herramienta fundamental para el procesamiento de datos secuenciales. A diferencia de las redes neuronales tradicionales *feedforward*, en las que la información fluye en una única dirección desde la entrada hasta la salida, las RNN incorporan conexiones recurrentes que permiten mantener un estado interno, proporcionando a la red la capacidad de "recordar" información de pasos anteriores en una secuencia. Esta característica hace que las RNN sean especialmente útiles en aplicaciones como el procesamiento de lenguaje natural, el reconocimiento de voz y la predicción de series temporales.

Con diferencia, la tarea más común relacionada con las series temporales es el *forecasting*: predecir qué sucederá a continuación en una serie. En este informe se busca predecir la temperatura dentro de 24 horas, dada una serie temporal de mediciones horarias de magnitudes como la presión atmosférica y la humedad, registradas por un conjunto de sensores en el tejado de un edificio. También, se analizará el funcionamiento de las RNN, sus diferencias con las redes feedforward.

2 EXPERIMENTAL DEVELOPMENT

En esta sección presentaremos los dos modelos utilizados, el tratamiento de los datos, y los parámetros a reportar.

2.1 Dataset

El trabajo se llevó acabo con un conjunto de datos meteorológicos de series temporales registrados en la estación meteorológica del Instituto Max Planck de Biogeoquímica en Jena, Alemania. En este conjunto de datos, se registraron 14 magnitudes diferentes (como temperatura, presión, humedad, dirección del viento, etc.) cada 10 minutos durante varios años. Los datos originales datan de 2003, pero el subconjunto de datos que descargaremos se limita al período 2009-2016.[1]

Variable	Descripción
p (mbar)	Presión atmosférica en milibares
T (°C)	Temperatura en grados centígrados
T_{pot} (K)	Temperatura potencial en kelvin
T_{dew} (°C)	Temperatura del punto de rocío en grados centígrados
rh (%)	Humedad relativa en porcentaje
VP_{max} (mbar)	Presión de vapor máxima en milibares
VP_{act} (mbar)	Presión de vapor actual en milibares
H_2O_C (mmol/mol)	Concentración de vapor de agua en milimoles por mol

Table 1. Algunas variables de interés y sus descripciones.

Todas las variables fueron descriptas en función del tiempo y normalizadas. Posteriormente se dividieron los datos en un 60% para el entrenamiento de la red, 20% para la validación y otros 20% para el testeo. En la Fig.1 se pueden observar algunas variables a lo largo del tiempo.

2.2 Modelos

- En primer lugar se entreno una red recurrente simple de 50 capas con una función de activación *Relu*, solamente utilizando los datos de temperatura.
- En segundo lugar se entro la misma red con todas las 14 variables meteorológicas disponibles.
- Luego, se analizaron la importancia de cada variable en la predicción. A partir de esta información se descartaron aquellas variables que perjudicaban al modelo.

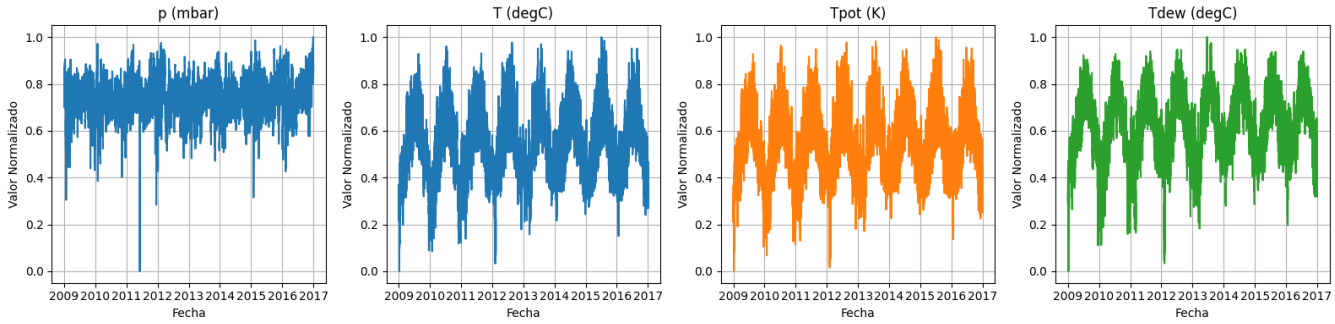

Figure 1. Cuatro variables arbitrarias del dataset en función del tiempo.

• Por último, se entreno un clasificador lineal con el objetivo de utilizarlo como un *benchmark* para comparar el modelo de RNN propuesto.

3 RESULTS AND CONCLUSIONS

El modelo unidimensional, aquel que se entrenó solo con la temperatura reportó un RMSE $=$ 0.22 y MAE $=$ 0.14 relativamente mejor en comparación al modelo de varias dimensiones que reportó un RMSE $= 0.36$ y MAE $= 0.34$

Al calcular la importancia de la variable, como la diferencia entre el error después de la permutación y el error base, se descartaron las variables *p (mbar)* y *rh (%)*. Esto se observa en la Fig.2.

Se entrenó nuevamente la red con las variables descartadas y se obtuvieron un RMSE $=$ 0.21 y MAE $=$ 0.2. Una relación de mejora del 58% para el RMSE, y del 59% para el MAE respecto al anterior.

Para el modelo de regresión lineal se reporta un RMSE $= 2.63$ y MAE $= 0.71$, lo que representa una peor predicción al problema esperado en comparación con las redes recurrentes propuestas previamente.

Finalmente se concluye que aún siendo una serie temporal no muy compleja o de poca variabilidad estructural, las redes recurrentes muestran un desempeño significativamente superior a un *benchmark* como lo es una red *feedforward* de regresión lineal. Como comentario final, se agrega que también se probaron modelos más avanzados como LSTM, los cuáles no presentaban diferencias con las RNN; se estima que se debe a la poca variabilidad estructural nuevamente.

1. *Dataset*. https://www.bgc-jena.mpg.de/wetter/

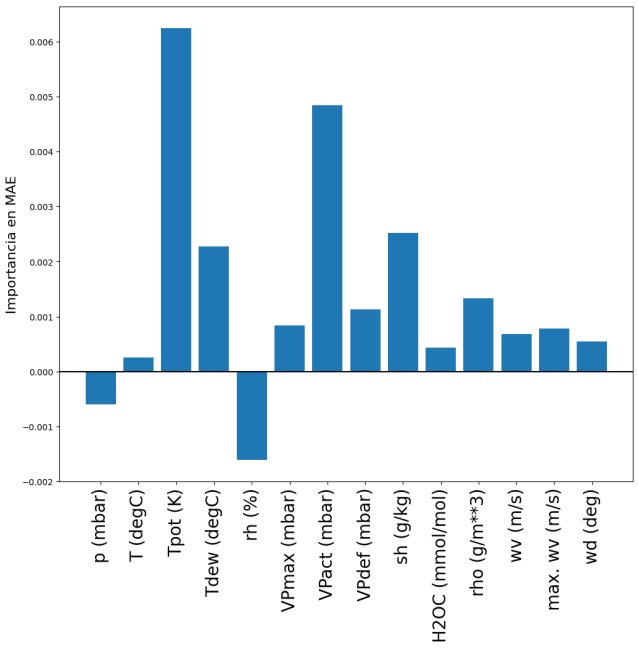

Figure 2. Importancia de las variables

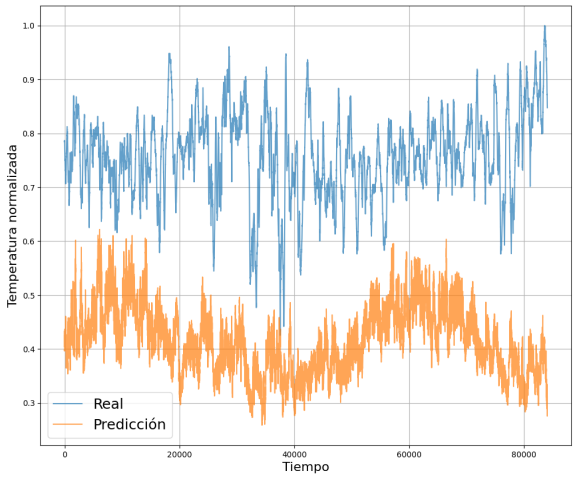

Figure 3. Gráfico de la temperatura real y la predicha por el modelo de varias variables sin dropout.