



Revista EIA, ISSN 1794-1237 /
e-ISSN 2463-0950
Año XVII/ Volumen 17/ Edición N.34
Julio-Diciembre de 2020
Reia34008 pág 1-16

Publicación científica semestral
Universidad EIA, Envigado, Colombia

**PARA CITAR ESTE ARTÍCULO /
TO REFERENCE THIS ARTICLE /**

Muñoz Herrera, W.; Bedoya, O.F.;
Rincón, M.E. (2020). Aplicación
de redes neuronales para la
reconstrucción de series de tiempo
de precipitación y temperatura
utilizando información satelital.
Revista EIA, 17(34), Julio-Diciembre,
Reia34008. [https://doi.org/10.24050/
reia.v17i34.1292](https://doi.org/10.24050/reia.v17i34.1292)

✉ *Autor de correspondencia:*

Muñoz Herrera, W. (Wilmer):
Universidad del Valle, Calle 13
Nº 100-00, edificio 331, espacio
2124, Cali, Colombia. Teléfono:
3105132630. Correo electrónico:
wilmer.munoz@correounivalle.edu.co

Recibido: 14-02-2019
Aceptado: 18-06-2020
Disponible online: 09-10-2020

Aplicación de redes neuronales para la reconstrucción de series de tiempo de precipitación y temperatura utilizando información satelital

✉ **WILMER MUÑOZ HERRERA¹**
OSCAR FERNANDO BEDOYA¹
MAURICIO EDILBERTO RINCÓN¹

1. Universidad del Valle

Resumen

Las técnicas de inteligencia artificial como las redes neuronales artificiales (RNA) permiten resolver una gran variedad de problemas relacionados con diferentes áreas del conocimiento tales como la medicina, la Bioinformática e incluso las telecomunicaciones. En muchos casos, las redes neuronales se utilizan para predecir el comportamiento de una variable con base en datos históricos previos y en un conjunto de variables predictoras. En este artículo se aborda el problema particular de la reconstrucción de información faltante de las estaciones meteorológicas utilizando RNAs. La falta de este tipo de información afecta principalmente los estudios climáticos en los que se utiliza información meteorológica. Estos estudios pueden permitir evitar las amenazas significativas en el desarrollo sustentable de nuestra sociedad, los recursos naturales, especies y la misma vida del ser humano. En este artículo se proponen modelos basados en redes neuronales artificiales e información satelital para el llenado de datos faltantes en las estaciones meteorológicas y reconstrucción espacial de las variables de precipitación y temperatura para la región de Departamento de Valle del Cauca, Colombia. Los resultados obtenidos alcanzan los coeficientes de correlación de alrededor de 0.9, con errores más pronunciados en cerca de 50 mm/mes en precipitación y 2 °C en temperatura.

Palabras Claves: Redes Neuronales, Llenado de Datos Faltantes, Sensores Remotos, Series de Tiempo.

Application of neural networks for the reconstruction of time series of precipitation and temperature using satellite information

Abstract

Artificial intelligence techniques such as artificial neural networks (ANN) allow solving a wide variety of problems related to different areas of knowledge such as medicine, Bioinformatics and even telecommunications. In many cases, neural networks are used to predict the behavior of a variable based on previous historical data and a set of predictor variables. This article deals with the particular problem of the reconstruction of missing information from meteorological stations using ANNs. The lack of this type of information mainly affects climate studies in which meteorological information is used. These studies can make it possible to avoid significant threats to the sustainable development of our society, natural resources, species and the very life of the human being. This article proposes models based on artificial neural networks and satellite information for the filling of missing data in meteorological stations and spatial reconstruction of the variables of precipitation and temperature for the Valle del Cauca, Colombia. The results obtained reach the correlation coefficients of around 0.9, with more pronounced errors in about 50 mm/month in precipitation and 2 °C in temperature.

Keywords: Neural Networks, Data Filling, Remote Sensors, Time Series.

1. Introducción

El conocimiento del comportamiento ambiental de un territorio se basa en los datos característicos de precipitación y temperatura. Su estudio se fundamenta en las series temporales, las cuales no siempre están disponibles o están incompletas. La modelación de las variables ambientales permite su entendimiento y la generación del pronóstico para la planificación territorial, agricultura, ecología, análisis y la conservación de los recursos naturales (Nayak *et al.*, 2004; Nourani, Alami and Aminfar, 2009; Andreadis *et al.*, 2017). La investigación de estas variables ambientales cumplen un papel importante para la mitigación de los impactos presentados por las sequías (Nourani, Alami and Aminfar, 2009; Son *et al.*, 2012)(Vancutsem *et al.*, 2010).

Para obtener una modelación eficiente de los procesos ambientales es necesario disponer de una fuente de información completa y precisa sobre una base temporal y/o espacial (Jeffrey *et al.*, 2001). Las inconsistencias o debilidad de los datos meteorológicos pueden atribuirse a los cambios de ubicación, calidad de equipos, técnicas de observación, periodos discretos, periodos intermitentes sin registros y errores sistemáticos (Peck, 1997; Jeffrey *et al.*, 2001). La existencia de valores atípicos en los registros de observación puede asumirse como falta de información, debido a que deben ser despreciados. Los datos atípicos o erróneos pueden dañar la salida de un modelo numérico (Jeffrey *et al.*, 2001). De esta manera, los problemas de modelación climática pueden verse reflejados por la falta de información temporal y espacial. La necesidad de lograr obtener información espacial y temporal detallada ha llevado a los investigadores a utilizar métodos satelitales para llenar los vacíos (Vancutsem *et al.*, 2010).

Las tecnologías de detección remota por satélite brindan la oportunidad de mejorar la calidad de los datos de entrada a los modelos hidrológicos, en particular el conjunto de datos de precipitación CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) que dispone información de alta resolución combinada por información satelital e información de estaciones meteorológicas, in situ (Le *et al.*, 2017). De esta misma forma los productos derivados de los sensores del espectrorradiómetro de las imágenes satelitales de MODIS (Moderate-Resolution Imaging Spectroradiometer) permiten obtener las temperaturas superficiales (Vancutsem *et al.*, 2010).

Las herramientas clásicas para la modelación de las series temporales, son básicamente modelos lineales que suponen que los datos son estacionarios (Nourani, Alami and Aminfar, 2009). Como también, el uso de métodos ponderados, modelos determinísticos y estocásticos (Hingray and Ben Haha, 2005; Teegavarapu and Chandramouli, 2005). El análisis de estas series de tiempo requiere mapear las relaciones complejas entre los valores de entrada y salida, ya que los valores de predicción son calculados en función de los patrones identificados en los datos de observación en campo (Nayak *et al.*, 2004).

Las redes neuronales artificiales han demostrado tener una gran capacidad para pronosticar y modelar las series de tiempo hidrológicas no lineales, siendo una herramienta adecuada para los problemas de modelado en series temporales, en el que ofrece un enfoque eficaz para el manejo de grandes cantidades de datos dinámicos, no lineales y con presencia de ruidos (Nayak *et al.*, 2004; Nourani, Alami and Aminfar, 2009). Sin embargo, no se ha realizado la reconstrucción de información con la ayuda de información satelital. La hipótesis que se plantea en esta investigación es que al usar la información satelital se puede lograr obtener resultados de reconstrucción de series de tiempo y espacial de la información meteorológica con mayor precisión.

En este artículo se presenta la construcción de información completa en términos espaciales y temporales de las variables de precipitación y temperatura basado en información de estaciones meteorológicas e imágenes satelitales, con el apoyo de la integración de redes neuronales artificiales. El resto de este artículo está dividido en secciones. En la sección 2 se presenta la metodología utilizada para obtener los modelos. En la sección 3 los resultados obtenidos por parte de los modelos. Finalmente, en la sección 4 se presenta las conclusiones.

2. Materiales y Métodos

En esta sección se propone un modelo basado en redes neuronales y datos satelitales para la reconstrucción de información meteorológica faltante para la región del Valle del Cauca. Cada uno de los pasos de la metodología utilizada se detalla a continuación.

2.1. Zona de Estudio

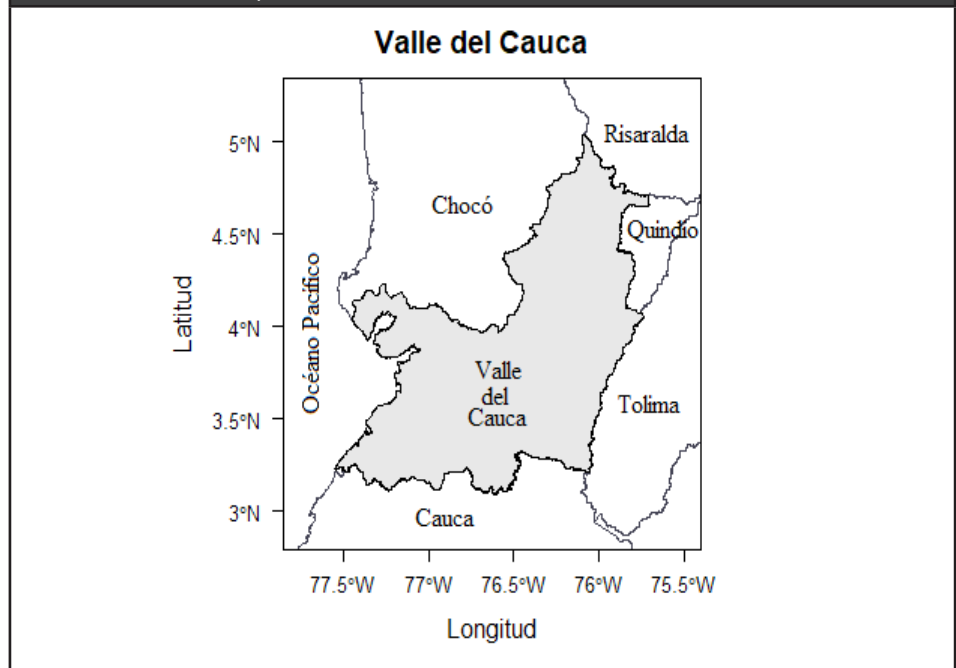
La zona de estudio es el departamento del Valle del Cauca que se encuentra ubicada en la región pacífica y andina de Colombia, limitando con los departamentos de Chocó, Risaralda, Quindío, Tolima, Cauca y el Océano Pacífico (**Figura 1**). La región de estudio está comprendida en las coordenadas geográficas 03°05'40"N, 77°32'52"W y 05°02'16"N, 75°42'21"W y con una extensión superficial alrededor de 22.000 Km² de área. La precipitación promedio anual en la parte media del Valle geográfico del Río Cauca en su recorrido por el departamento, desde Puerto Tejada hasta Cartago no supera los 1500 mm/año, con excepción de Cali, Yumbo, Cerrito y Candelaria donde oscila alrededor de 1000 mm/año, en la parte alta de la vertientes la precipitación tiene valores cercanos a los 2000 mm/año, en la parte baja oscila los 4000 mm/año en

las estribaciones de las cordilleras, 1500 mm/año en la desembocadura del río Cauca y en la parte pacífica puede alcanzar los 6000 mm/año (Pabón, Daniel;Caicedo, 2001).

La temperatura en Valle del Río Cauca presenta dos épocas cálidas y frías. La primera época cálida abarca los meses de febrero y marzo con valores de temperatura media mensual entre 0.5 °C y 0.7 °C por encima de la media anual. La segunda época cálida es menos definida que la primera y abarca los meses de julio y agosto. La primera época fría corresponde a la transición del mes de Abril hasta Junio con temperaturas entre 0.2 °C y 0.3 °C por debajo de la media anual. La segunda época fría se considera más alta para los meses de octubre a diciembre, los cuales presentan temperaturas de 0.3 °C y 0.8 °C por debajo de la media anual (Pabón, Daniel;Caicedo, 2001).

Figura 1. Mapa del Departamento del Valle del Cauca.

Fuente: Elaboración Propia.



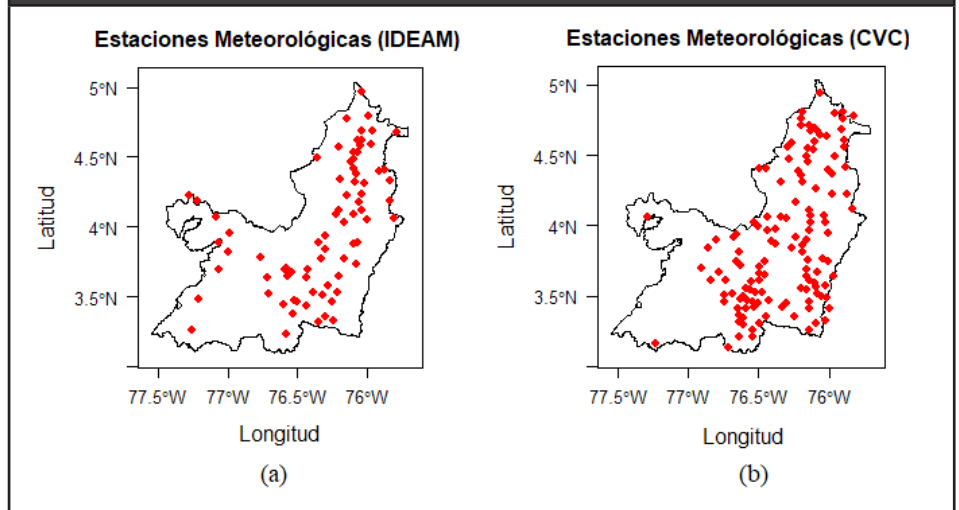
2.2. Información Meteorológica Local

La información histórica de las estaciones para la precipitación y temperatura en la región del Valle del Cauca se obtuvo de las estaciones operadas por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), la cual corresponde a 79 estaciones que se encuentran distribuidas por toda la región, como se muestra en la **Figura 2a**. Además, información de precipitación y temperatura de estaciones por parte de la corporación autónoma del Valle del Cauca (CVC), el cual tiene 131 estaciones distribuidas en toda la región del Valle del Cauca, como se muestra en la **Figura 2b**.

La información meteorológica tiene algunos vacíos en ciertos periodos, esto se debe a que las estaciones no capturaron información por fallas, mantenimiento o fuera de servicio, como también puede presentar información atípica. Por tanto, es necesario realizar un tratamiento de la información, mediante un análisis exploratorio de datos.

Figura 2. (a). Estaciones meteorológicas por parte del IDEAM. (b). Estaciones meteorológicas por parte de la CVC.

Fuente: Elaboración Propia.

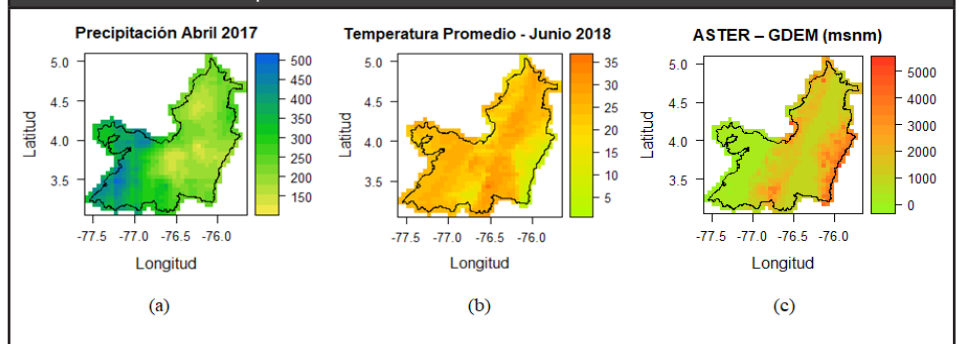


2.3. Información Satelital y Datos Climáticos Globales

Para el apoyo de la reconstrucción de la información meteorológica, se utiliza información de datos globales de CHIRPS para la precipitación (**Figura 3a**), el cual dispone de series históricas mensual desde 1981 hasta el presente, con resolución espacial de $0.05^\circ \times 0.05^\circ$ (5.6 Km \times 5.6 Km, aprox.) (Katsanos, Retalis and Michaelides, 2016). Además, el uso de la información satelital de MODIS LTS para la temperatura (**Figura 3b**), con resolución espacial de $0.05^\circ \times 0.05^\circ$ (5.6 Km \times 5.6 Km, aprox.) desde el 2000 hasta el presente (Wan, 2013). Por otro lado, también se hace uso de información satelital ASTER (**Figura 3c**), con resolución espacial de 1 arc-segundo (30 mts, aprox.) para la construcción de la topografía en la zona de estudio, cada valor del pixel corresponde a la altura en metros sobre el nivel del mar (Tachikawa *et al.*, 2011).

Figura 3. (a). Precipitación acumulada mensual desde CHIRPS (mm/mes). (b). Temperatura promedio mensual desde MODIS LTS ($^\circ\text{C}/\text{mes}$). (c). Modelo digital de elevaciones desde ASTER (msnm).

Fuente: Elaboración Propia.



2.4. Análisis Exploratorio de Datos Meteorológicos Locales

Antes de realizar cualquier tipo de análisis climático, es importante tener una visión general de las variables atmosféricas que se desea estudiar mediante diagramas de cajas y series, e identificar la existencia de valores atípicos que pueden presentarse

en los datos. Los valores atípicos se pueden presentar debido a que las observaciones en algunos casos son capturadas con instrumentos manuales y digitalizados por personas, lo cual implica la existencia de cometer algún tipo de error.

El proceso de detección de valores atípicos se utilizó el rango intercuartílico (IQR), es decir, la diferencia existente entre el primer cuartil Q_1 y el tercer cuartil Q_3 (Kolden and Weisberg, 2007; Homar *et al.*, 2010). En un diagrama de cajas se puede representar e identificar los tres cuartiles, el rango intercuartílico y los valores atípicos, en los valores atípicos se consideran dos tipos: (a) valor atípico leve (**Ecuación 1**) y (b) valor atípico extremo (**Ecuación 2**). A continuación, se presenta el cálculo de estos dos tipos de valores atípicos.

Valor atípico leve:

$$\begin{aligned} A_i &< Q_1 - 1.5 \times IQR \\ A_s &> Q_3 + 1.5 \times IQR \end{aligned} \quad (1)$$

Siendo A_i y A_s valores atípicos inferiores y superiores, Q_1 y Q_3 el primer y tercer cuartil, y IQR el rango intercuartílico ($Q_3 - Q_1$).

Valor atípico extremo:

$$\begin{aligned} A_i &< Q_1 - 3 \times IQR \\ A_s &> Q_3 + 3 \times IQR \end{aligned} \quad (2)$$

Siendo A_i y A_s valores atípicos inferiores y superiores, Q_1 y Q_3 el primer y tercer cuartil, y IQR el rango intercuartílico ($Q_3 - Q_1$).

Las series de precipitación y temperatura presentaron una baja existencia de valores atípicos extremos, 12% para la precipitación y 8% para la temperatura del conjunto de datos se consideraron valores atípicos extremos, los cuales fueron eliminados de la base de datos. Los valores atípicos leves se presentan con más frecuencia, sin embargo, no fueron eliminados debido a que estas variables pueden tener alteraciones que son asociados a ciertos fenómenos físicos que pueden ocurrir en el ambiente. Además, se conoce que algunas variables climáticas pueden presentar fluctuaciones, por lo cual se considera omitir los valores atípicos leves y eliminar los valores atípicos extremos.

2.5. Llenado de Datos Faltantes con Redes Neuronales

Una vez identificados y eliminados los valores atípicos extremos, se realiza el llenado y reconstrucción de datos faltantes para las series de precipitación total acumulada mensual y temperatura máxima y mínima promedio mensual, con ayuda de información satelital para el soporte de información faltante y redes neuronales artificiales como herramienta para la modelación de los datos.

Debido a que los datos de precipitación y temperatura presentan vacíos en algunos meses o periodos es importante reconstruir las series de tiempo para obtener un mejor resultado en los diferentes estudios climáticos, como también, la distribución espacial de las estaciones en la zona de estudio que no son homogéneas, es decir que, existen zonas sin información puntual. Es fundamental tener una información adicional que permita mitigar esta problemática de información, de esta manera, se hace uso de información satelital o datos climáticos globales que tienen una distribución espacial homogénea y periodos de información completa. Para la precipitación se hace uso de los datos CHIRPS y para las temperaturas el uso de productos MODIS, en específico el producto MOD11C3. Estos dos tipos de información adicional tienen una resolución espacial de $0.05^\circ \times 0.05^\circ$ (5.6 Km \times 5.6 Km, aprox.).

En este artículo se propone usar una red neuronal para predecir cada una de las variables climáticas. Cada red tiene una cantidad de neuronas de entrada que corresponden a las variables que determinan la variable climática a predecir. El objetivo de cada red neuronal es predecir ya sea la precipitación o la temperatura en cada coordenada, con separación de 5.6 Km × 5.6 Km. Para predecir la variable de precipitación, se utilizan seis neuronas de entrada: Longitud, Latitud, Elevación, Año, Mes y CHIRPS. Para predecir la variable de temperatura, seis neuronas de entrada: Longitud, Latitud, Elevación, Año, Mes y MODIS. En las capas ocultas, el número de neuronas en los dos tipos de variables climáticas se realiza mediante el proceso de ensayo-error y con la ayuda de algunas heurísticas o reglas (Masters, 1993; Wang, 1994; Lenard, Alam and Madey, 1995)

Finalmente, en la capa de salida se obtiene sólo una neurona denominada el valor de precipitación o temperatura de forma puntual cada 5.6 Km aproximadamente de distribución espacial. En la **Figura 4** se muestra un ejemplo del proceso de entrenamiento de la variable precipitación. La imagen de la parte izquierda representa los datos de CHIRPS o datos satelitales y la imagen derecha los datos capturados por las 210 estaciones meteorológicas. En la **Figura 5** se presenta el resultado de predicción para el llenado de datos faltantes en las series de tiempo y reconstrucción de información en zonas no existentes. De esta misma manera se realiza el proceso de aprendizaje y predicción para la variable de temperatura, con la ayuda del producto MODIS – LTS.

Figura 4. Diagrama del proceso de aprendizaje de la Red Neuronal Artificial (Precipitación Mensual).

Fuente: Elaboración Propia.

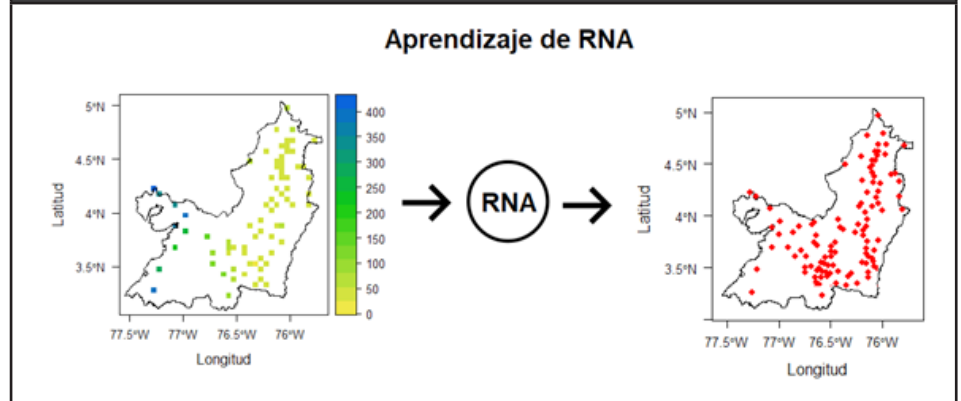
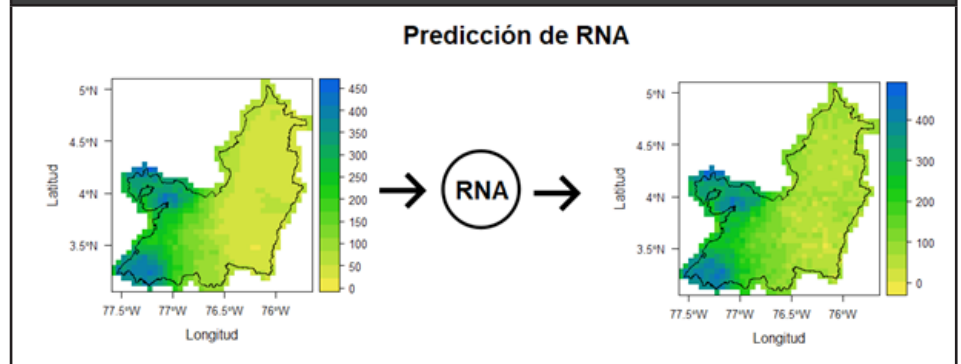


Figura 5. Diagrama del proceso de predicción de la Red Neuronal Artificial (Precipitación mensual).

Fuente: Elaboración Propia.



El tipo de red utilizada para el proceso de reconstrucción de información, corresponde a la red perceptrón multicapa con el algoritmo de aprendizaje *Backpropagation* y el uso de la función de activación *sigmoide*. Siguiendo las sugerencias por parte de las heurísticas o reglas se construyen cuatro posibles redes neuronales para las dos variables climáticas que tienen seis neuronas de entrada y una neurona de salida, pero la cantidad de capas ocultas y neuronas varían. Estas topologías de red se representan numéricamente como 6-3-1, 6-3-2-1, 6-4-3-2-1 y 6-5-4-3-2-1. Por ejemplo, en la topología 6-3-2-1, se tiene seis neuronas de entrada, dos capas ocultas, una de tres neuronas y otra de dos, y finalmente una neurona de salida. En las **Figuras 6 y 7** se ilustran las topologías de las redes neuronales propuestas.

Figura 6. (a). Estructura del modelo de red neuronal 6-3-1 para precipitación y temperatura. (b). Estructura del modelo de red neuronal 6-3-2-1 para precipitación y temperatura.

Fuente: Elaboración Propia.

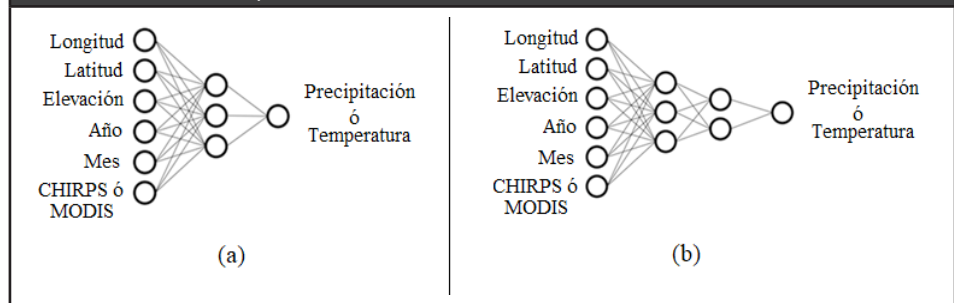
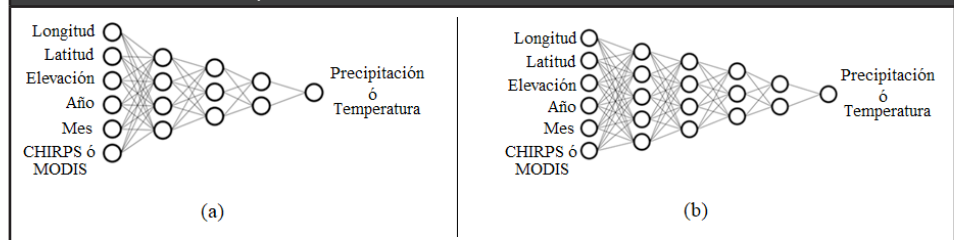


Figura 7. (a). Estructura del modelo de red neuronal 6-4-3-2-1 para precipitación y temperatura. (b). Estructura del modelo de red neuronal 6-5-4-3-2-1 para precipitación y temperatura.

Fuente: Elaboración Propia.



Además, para evitar la sobrestimación de los datos en el proceso de aprendizaje es necesario realizar la normalización o escalado de los datos de entrada en el rango [0-1], ya que existen variables que pueden oscilar en rangos de 3.000, 1.500, 30, 12 o 5 aproximadamente. Existen varias formas de efectuar el proceso de normalización, en este caso se realiza mediante la **Ecuación 3**.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Donde, X' corresponde al nuevo valor escalado en el rango [0,1], X el valor inicial, X_{min} el valor mínimo del conjunto de datos y X_{max} el valor máximo del conjunto de datos.

La independencia entre los conjuntos de datos es muy importante. Para realizar el entrenamiento y prueba de las redes neuronales es necesario realizar la validación cruzada. Para esto, se divide el conjunto de 52252 registros de precipitación y 3634 registros de temperatura en diez subconjuntos de datos de forma aleatoria de los cuales nueve pertenecen al conjunto de datos de entrenamiento y uno al conjunto de prueba,

cada uno de diez subconjuntos deben pertenecer a prueba y los otros como conjunto de entrenamiento, de esta manera se obtiene 10 experimentos. Cada uno de estos experimentos se realizó mediante la ayuda de la librería de *neuralnet* en el software R CRAN.

Para la evaluación del desempeño y rendimiento de cada uno de los ensayos, se dispone de mediciones estadísticas de error basadas en el conjunto de datos observados y calculados. Las mediciones de error utilizadas son: Coeficiente de Pearson, Error Medio Cuadrático, Error Medio Absoluto y Raíz del Error Medio Cuadrático. Otra herramienta que permite evaluar los modelos es el Diagrama de Taylor, el cual compara varios modelos y determina cuál de ellos es más predictivo, basándose, en algunas de las mediciones de error mencionados anteriormente.

Una vez identificadas las estructuras de las redes para precipitación y temperatura que obtengan el menor error de incertidumbre, se procede a la reconstrucción espacial y llenado de datos faltantes en las series de tiempo.

3. Resultados

3.1. Medidas de error

Existen algunas herramientas o métricas para la evaluación del desempeño de predicción de cada uno de los métodos y comparación entre los métodos de aprendizaje automático presentados en esta investigación. Las herramientas más utilizadas son: mediciones de error estadístico, validación cruzada y diagrama de Taylor. Además, es importante resaltar que las variables de estudio están distribuidas espacialmente, lo cual implica que las mediciones de error deben ser diferentes de acuerdo a su zona o región.

Error Medio Absoluto (MAE)

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (4)$$

Donde, n corresponde a la cantidad de registros, y_i es el valor de predicción y x_i corresponde al valor verdadero.

Error Medio Cuadrático (MSE)

$$MSE = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n} \quad (5)$$

Donde, n corresponde a la cantidad de registros, y_i es el valor de predicción y x_i corresponde al valor verdadero.

Raíz del Error Medio Cuadrático (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}} \quad (6)$$

Donde, n corresponde a la cantidad de registros, y_i es el valor de predicción y x_i corresponde al valor verdadero.

Coeficiente de Correlación de Pearson (R)

$$r_{xy} = \frac{n \cdot \sum x_i y_i - \sum x_i \cdot \sum y_i}{\sqrt{n \cdot \sum x_i^2 - (\sum x_i)^2} \sqrt{n \cdot \sum y_i^2 - (\sum y_i)^2}} \quad (7)$$

Donde, n corresponde a la cantidad de registros, y_i es el valor de predicción y x_i corresponde al valor verdadero.

Coefficiente de determinación (R^2)

$$R^2 = (r_{xy})^2 \tag{8}$$

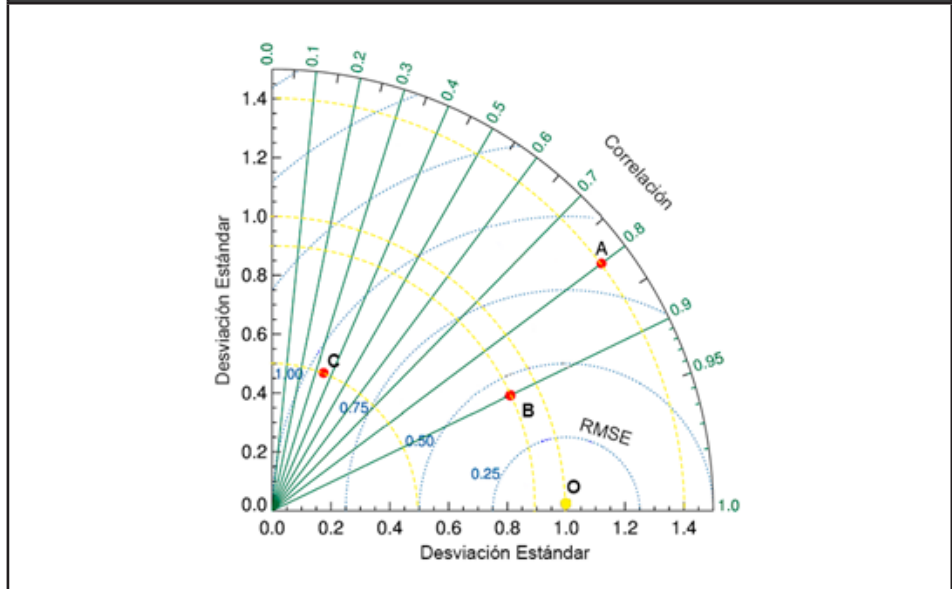
Donde, r_{xy} corresponde al coeficiente de correlación.

Diagrama de Taylor

Taylor en el 2001 propuso un tipo de diagrama que facilita el grado de correspondencia de los valores observados y modelados a través de varias técnicas como se muestra en la **Figura 8**. Este diagrama permite identificar la relación o distancia de similitud que existe entre los datos modelados y observados. El diagrama es muy utilizado para la evaluación de modelos climáticos y otros aspectos medio ambientales, los cuales relaciona los tres aspectos estadísticos de los datos, raíz del error medio cuadrático (RMSE), desviación estándar (sd) y coeficiente de Correlación de Pearson (r) (Gleckler, Taylor and Doutriaux, 2008; Kawase *et al.*, 2009). El diagrama se representa como se muestra en la **Figura 8**, donde la desviación estándar se ubica en las abscisas (x, y) y el arco de color *amarillo*, el coeficiente de correlación en las líneas inclinadas de color *verde*, la raíz del error medio cuadrático corresponde a los arcos de color *azul*, y los puntos de color *amarillo* y *rojo* representan los valores de observación y modelados.

Figura 8. Diagrama de Taylor.

Fuente: Elaboración Propia.



3.2. Resultados Obtenidos

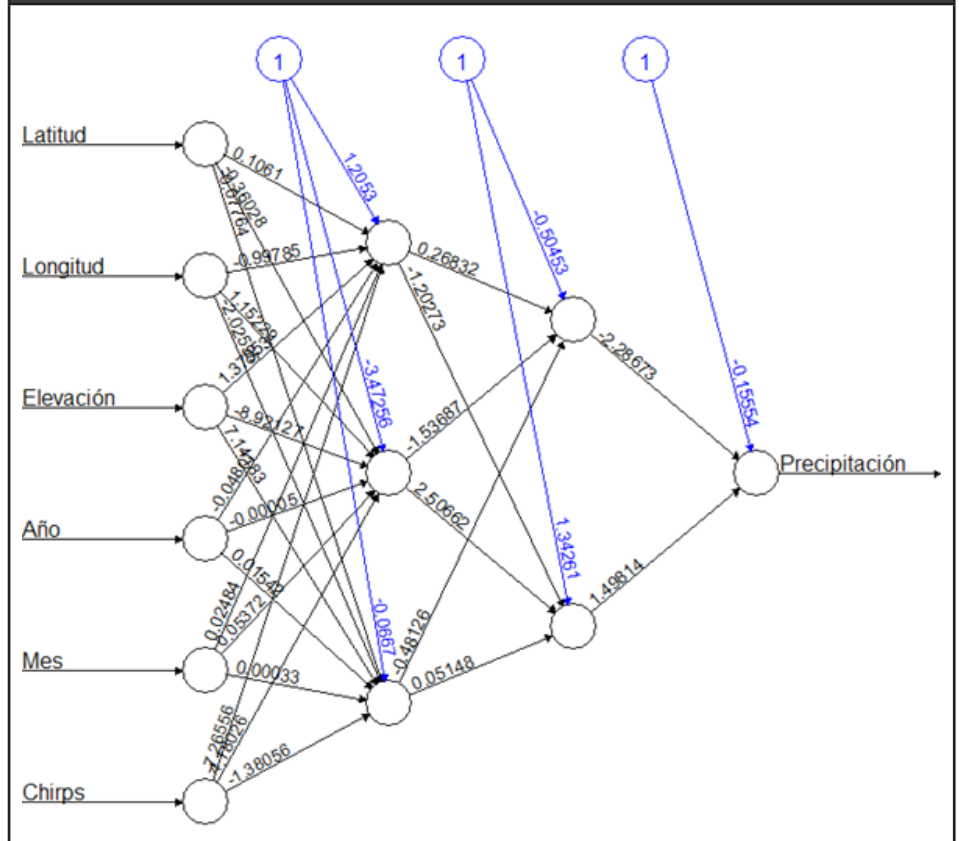
La calidad de los datos de entrada para los diferentes estudios climáticos es importante para la obtención de buenos resultados, de esta manera los datos de precipitación y temperatura de las estaciones meteorológicas cumplen un papel muy significativo. Por tanto, es necesario realizar un análisis exploratorio general de los datos, eliminación de datos atípicos, llenado de datos faltantes y reconstrucción de información.

La construcción en zonas no existentes y reconstrucción temporal de la información se obtuvo con la ayuda de información satelital, datos globales, modelo digital de terreno y técnica de redes neuronales. Para predecir las variables de precipitación y temperatura se probaron las cuatro topologías de red neuronal que se indicaron anteriormente, es decir, las topologías 6-3-1, 6-3-2-1, 6-4-3-2-1 y 6-5-4-3-2-1.

Además, en cada uno de estos experimentos se utiliza la función de activación sigmoide. En la **Figura 9** se muestra el ejemplo de la red neuronal para la topología 6-3-2-1 cuando se pretende predecir la precipitación.

Figura 9. Red neuronal artificial para precipitación usando la topología 6-3-2-1.

Fuente: Elaboración Propia.



Los resultados estadísticos para las pruebas de las cuatro topologías para la precipitación y temperatura se muestran en las **Figuras 10 y 11**. Por ejemplo, en la **Figura 10a** se muestra el error medio absoluto (MAE) para las cuatro topologías (RN1, RN2, RN3, RN4) en el eje X y los 10 experimentos en el eje Y. La escala de color verde a rojo representa los valores de 34mm/mes a 36mm/mes, donde se puede observar que la topología RN4 tiene una menor diferencia entre los valores reales y modelados, respecto a las otras topologías. Sin embargo, las diferencias de las topologías RN1, RN2 y RN3 son similares a RN4. De igual manera, la **Figuras 10b** representa la raíz del error medio cuadrático (RMSE). Además, los niveles de coeficientes de correlación en las **Figuras 10c y 10d** están representados en escala de color de rojo para una correlación de 0 y verde para 1, evidenciando que los resultados estadísticos tienen una relación entre sí. Por otro lado, en la **Tabla 1** se presenta el resumen de los valores promedios estadísticos de los diez experimentos para las cuatro topologías.

Figura 10. Mediciones estadísticas de las cuatro topologías de redes neuronales para los diez subconjuntos de datos en precipitación, (a). Error Absoluto, (b). Raíz Error Medio Cuadrático, (c). Coeficiente de Correlación de Pearson, (d). Coeficiente de determinación.

Fuente: Elaboración Propia.

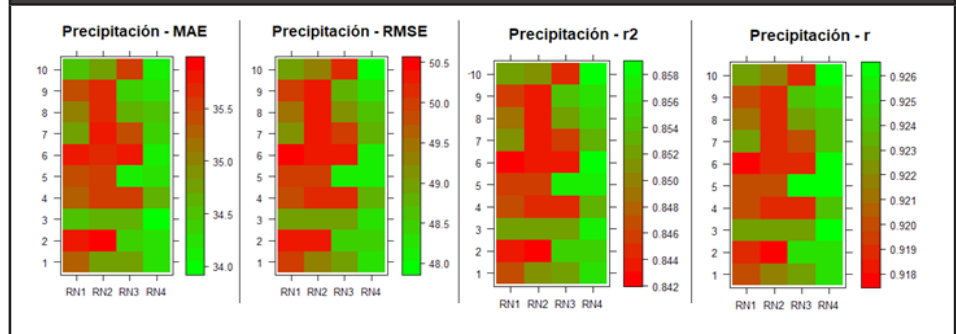


Figura 11. Mediciones estadísticas de las cuatro topologías de redes neuronales para los diez subconjuntos de datos en temperatura, (a). Error Medio Absoluto, (b). Raíz Error Medio Cuadrático, (c). Coeficiente de Correlación de Pearson, (d). Coeficiente de determinación.

Fuente: Elaboración Propia.

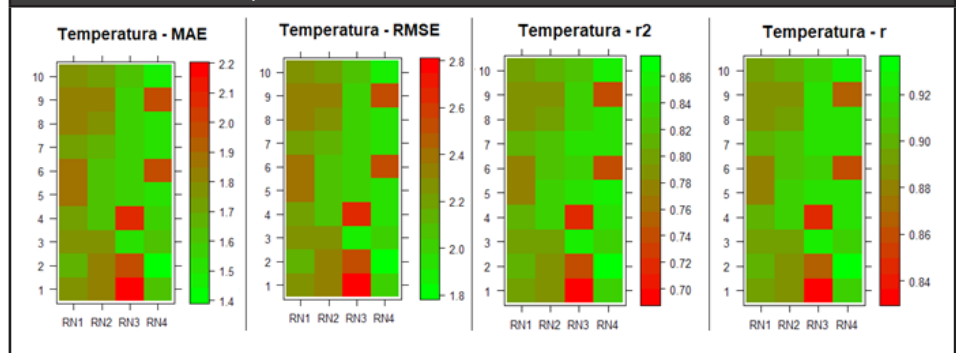


TABLA 1. RESULTADOS PROMEDIOS ESTADÍSTICOS DE LOS DIEZ EXPERIMENTOS PARA LA PRECIPITACIÓN Y TEMPERATURA DE LAS CUATRO TOPOLOGÍAS.

FUENTE: ELABORACIÓN PROPIA.

VARIABLE	TOPOLOGÍA	MAE	MSE	RMSE	R	R ²
Precipitación	6-3-1	35,188	2471	49,696	0,921	0,848
	6-3-2-1	35,473	2500	49,999	0,920	0,846
	6-4-3-2-1	34,840	2422	49,209	0,922	0,851
	6-5-4-3-2-1	34,260	2344	48,417	0,925	0,855
Temperatura	6-3-1	1,785	5,134	2,265	0,892	0,795
	6-3-2-1	1,729	4,853	2,201	0,898	0,807
	6-4-3-2-1	1,717	4,903	2,193	0,896	0,805
	6-5-4-3-2-1	1,618	4,329	2,067	0,909	0,827

Teniendo en cuenta los resultados obtenidos en las mediciones de error estadísticos de las cuatro topologías y cada uno de los experimentos para la variable de precipitación, se puede decir que los resultados de los experimentos en cada una de las topologías son muy similares, lo cual implica que en el conjunto de datos no existe una sobreestimación de los subconjuntos que fueron seleccionados de forma aleatoria. El

error medio absoluto entre los valores modelados y reales está alrededor de los 35 mm/mes, con errores más pronunciados en 50 mm/mes de RMSE y un coeficiente de correlación alrededor de 0.92. Este error de 50 mm/mes es aceptable, ya que puede corresponder a un error de un día de lluvia en el mes.

Por su parte, comparando los promedios de los experimentos de las cuatro topologías se identifican que tienen una relación muy estrecha, de tal manera que, mientras más compleja o mayor cantidad de capas ocultas y neuronas en la red, los resultados son muy similares. Lo cual implica que no sería necesario complejizar la red para obtener buenos resultados.

Por otro lado, los resultados estadísticos de la variable de temperatura de los 10 experimentos en las topologías 6-4-3-2-1 y 6-5-4-3-2-1 presentan variaciones notables. Sin embargo, en las topologías 6-3-1 y 6-3-2-1 las variaciones son menores, es decir, se puede considerar que son muy similares, ya que se obtiene un error medio absoluto alrededor de 1.7 °C, con errores más pronunciados de 2 °C en RMSE y con coeficiente de correlación cercano a 0.9. De esta manera, el proceso de validación cruzada permite identificar que las redes con mayor número de capas ocultas y neuronas presentan grandes variaciones entre los subconjuntos de datos. Además, con la ayuda del diagrama de Taylor se puede identificar las distancias que existen entre cada uno de los experimentos y los datos de observación, como se muestra en las **Figuras 12 a 15**.

Los diagramas de Taylor permiten evidenciar los resultados estadísticos de correlación y raíz del error medio cuadrático. Los puntos azules representan los valores de predicción en comparación con los valores reales. En las **Figuras 12 y 13** los puntos azules presentan separaciones mínimas entre los experimentos, de tal manera que no presentan sobrestimación en los subconjuntos, dado que el coeficiente de correlación está alrededor de 0.91 a 0.92 y valores RMSE de 50mm/día para la precipitación y para la temperatura un coeficiente de correlación de alrededor de 0.87 a 0.91 y valores RMSE de 2°C. De esta manera, las redes neuronales con topologías 6-3-1 y 6-3-2-1 permiten obtener resultados aceptables de predicción para la precipitación y temperatura. Además, las **Figuras 14 y 15** presentan resultados de correlación y valores de RMSE similares a las RN1 y RN2 de las precipitaciones, mientras que para la temperatura las diferencias de separación crecen, es decir, obtienen un coeficiente de correlación de alrededor de 0.82 a 0.92 con valores de RMSE de 2°C a 3°C. En general, el análisis de resultados a través de los diagramas de Taylor muestra que la precipitación presenta resultados similares en las cuatro topologías de red propuestas, con resultados aceptables. Sin embargo, para la temperatura solo las RN1 y RN2 presentan resultados aceptables.

Figura 12. Diagrama de Taylor de RN1 para precipitación y temperatura.

Fuente: Elaboración Propia

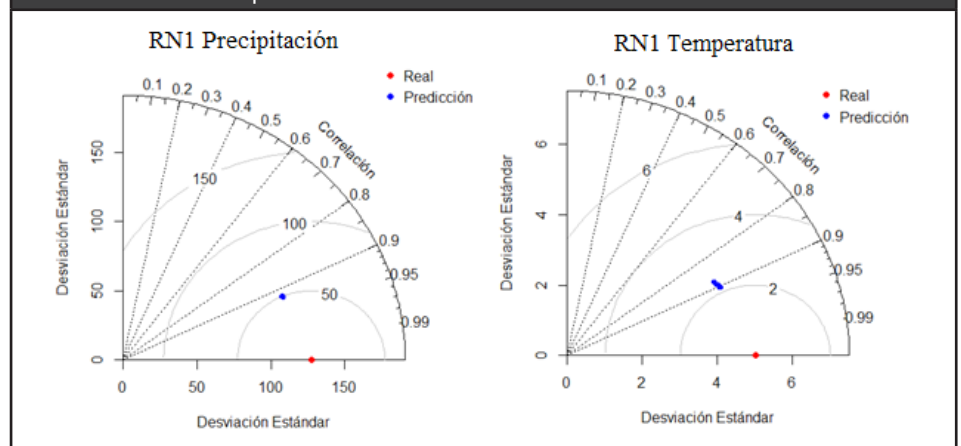


Figura 13. Diagrama de Taylor de RN2 para precipitación y temperatura.

Fuente: Elaboración Propia.

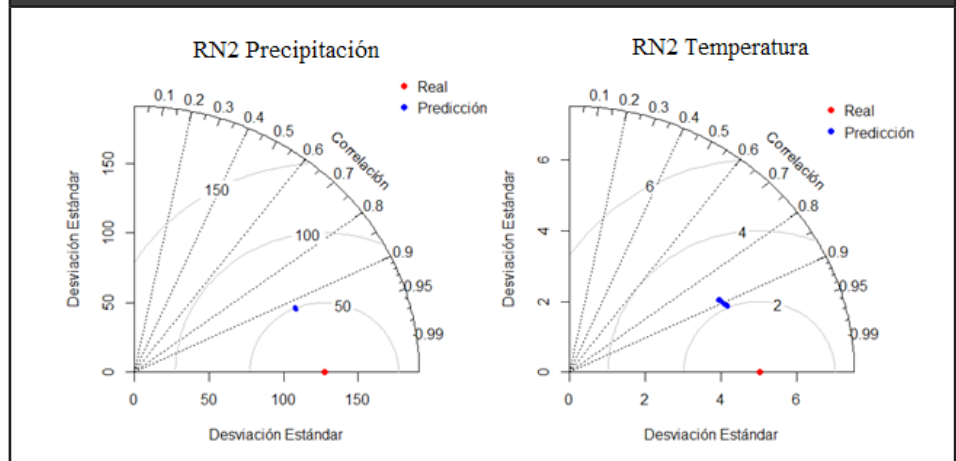


Figura 14. Diagrama de Taylor de RN3 para precipitación y temperatura.

Fuente: Elaboración Propia.

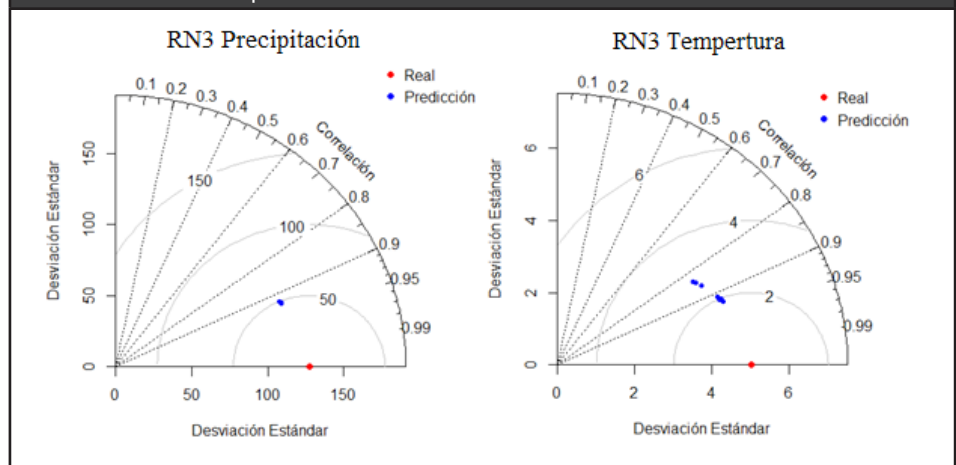
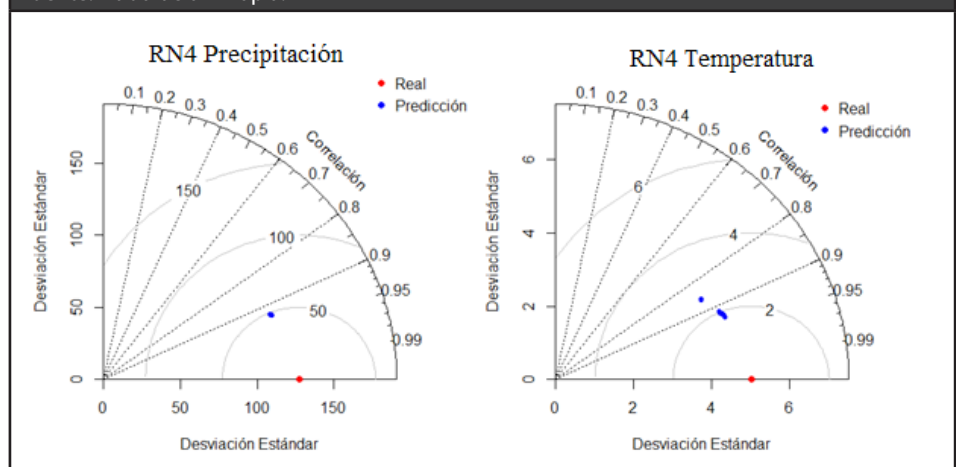


Figura 15. Diagrama de Taylor de RN4 para precipitación y temperatura.

Fuente: Elaboración Propia.



4. Conclusiones

En este artículo se logró comprobar que la información satelital de temperatura de MODIS y datos globales de precipitación CHIRPS permiten reconstruir información espacial y series de tiempo no existentes por parte de las estaciones meteorológicas locales. En particular, se utilizaron redes neuronales como método para la reconstrucción de series de tiempo climáticas, obteniendo niveles de correlación aceptables de alrededor de 0.92 para la precipitación y 0.9 para la temperatura, y con errores más pronunciados en cerca de 50 mm/mes en precipitación y 2 °C en temperatura. Esta información puede resultar indispensable para realizar proyecciones climáticas que abarquen toda la distribución espacial de la región de estudio.

Las redes neuronales artificiales de tipo perceptrón multicapa con el algoritmo de aprendizaje backpropagation y la función de activación sigmoide permite predecir los datos de precipitación y temperatura para la reconstrucción de datos faltantes. Sin embargo, dada la creciente evolución y amplia existencia de herramientas de inteligencia artificial se puede explorar otras técnicas, tales como: Deep Learning, Random Forest, Árboles de Decisión, Máquinas de Soporte Vectorial, Lógica Difusa, Algoritmos Genéticos, entre otras.

Agradecimientos

Agradecimiento por la colaboración al Grupo de Investigación GISMODEL, Centro de Investigación CIBioFi de la Universidad del Valle, Universidad del Valle, IDEAM, CVC, Chirps y MODIS. El reconocimiento al apoyo financiero por parte del Sistema General de Regalías (Fondo de Ciencia, Tecnología e Innovación), Gobernación del Valle del Cauca, bajo el Macroproyecto titulado "Implementación Centro Regional de Investigación e Innovación en Bioinformática y Fotónica, Cali, Valle del Cauca, Occidente" (BPIN 2013000100007).

Referencias

- Andreadis, K. M. *et al.* (2017) 'The Regional Hydrologic Extremes Assessment System: A software framework for hydrologic modeling and data assimilation', *PLOS ONE*. Edited by G. J.-P. Schumann. Public Library of Science, 12(5), p. e0176506. doi: 10.1371/journal.pone.0176506.
- Gleckler, P. J., Taylor, K. E. and Doutriaux, C. (2008) 'Performance metrics for climate models', *Journal of Geophysical Research Atmospheres*, 113(6), pp. 1–20. doi: 10.1029/2007JD008972.
- Hingray, B. and Ben Haha, M. (2005) 'Statistical performances of various deterministic and stochastic models for rainfall series disaggregation', *Atmospheric Research*. Elsevier, 77(1–4), pp. 152–175. doi: 10.1016/J.ATMOSRES.2004.10.023.
- Homar, V. *et al.* (2010) 'Recent trends in temperature and precipitation over the Balearic Islands (Spain)', *Climatic Change*, 98, pp. 199–211. doi: 10.1007/s10584-009-9664-5.
- Jeffrey, S. J. *et al.* (2001) 'Using spatial interpolation to construct a comprehensive archive of Australian climate data', *Environmental Modelling & Software*. Elsevier, 16(4), pp. 309–330. doi: 10.1016/S1364-8152(01)00008-1.
- Katsanos, D., Retalis, A. and Michaelides, S. (2016) 'Validation of a high-resolution precipitation database (CHIRPS) over Cyprus for a 30-year period', *Atmospheric Research*. Elsevier, 169, pp. 459–464. doi: 10.1016/J.ATMOSRES.2015.05.015.
- Kawase, H. *et al.* (2009) 'Intermodel variability of future changes in the Baiu rainband estimated by the pseudo global warming downscaling method', *Journal of Geophysical Research*, 114(D24), p. D24110. doi: 10.1029/2009JD011803.
- Kolden, C. A. and Weisberg, P. J. (2007) 'Assessing Accuracy of Manually-mapped Wildfire Perimeters In Topographically Dissected Areas', *Fire Ecology*. Springer International

Publishing, 3(1), pp. 22–31. doi: 10.4996/fireecology.0301022.

- Le, A. *et al.* (2017) 'Increasing the Accuracy of Runoff and Streamflow Simulation in the Nzoia Basin, Western Kenya, through the Incorporation of Satellite-Derived CHIRPS Data', *Water*. Multidisciplinary Digital Publishing Institute, 9(2), p. 114. doi: 10.3390/w9020114.
- Lenard, M. J., Alam, P. and Madey, G. R. (1995) 'The Application of Neural Networks and a Qualitative Response Model to the Auditor's Going Concern Uncertainty Decision', *Decision Sciences*. Wiley/Blackwell (10.1111), 26(2), pp. 209–227. doi: 10.1111/j.1540-5915.1995.tb01426.x.
- Masters, T. (1993) *Practical neural network recipes in C++*.
- Nayak, P. *et al.* (2004) 'A neuro-fuzzy computing technique for modeling hydrological time series', *Journal of Hydrology*. Elsevier, 291(1–2), pp. 52–66. doi: 10.1016/J.JHYDROL.2003.12.010.
- Nourani, V., Alami, M. T. and Aminfar, M. H. (2009) 'A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation', *Engineering Applications of Artificial Intelligence*. Pergamon, 22(3), pp. 466–472. doi: 10.1016/J.ENGAPPAI.2008.09.003.
- Pabón, Daniel;Caicedo, J. (2001) *Generalidades de la distribución espacial y temporal de la temperatura del aire y de la precipitación en Colombia*.
- Peck, E. L. (1997) 'QUALITY OF HYDROMETEOROLOGICAL DATA IN COLD REGIONS', *Journal of the American Water Resources Association*. John Wiley & Sons, Ltd, 33(1), pp. 125–134. doi: 10.1111/j.1752-1688.1997.tb04089.x.
- Son, N. T. *et al.* (2012) 'Monitoring agricultural drought in the Lower Mekong Basin using MODIS NDVI and land surface temperature data', *International Journal of Applied Earth Observation and Geoinformation*. Elsevier, 18, pp. 417–427. doi: 10.1016/J.JAG.2012.03.014.
- Tachikawa, T. *et al.* (2011) 'CHARACTERISTICS OF ASTER GDEM VERSION 2 1 Earth Remote Sensing Data Analysis Center (ERSDAC), 2 Mitsubishi Material Techno Corp. 3 University of Tokyo', *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, (90 m), pp. 3657–3660. doi: 10.1109 / IGARSS.2011.6050017.
- Teegavarapu, R. S. V. and Chandramouli, V. (2005) 'Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records', *Journal of Hydrology*. Elsevier, 312(1–4), pp. 191–206. doi: 10.1016/J.JHYDROL.2005.02.015.
- Vancutsem, C. *et al.* (2010) 'Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa', *Remote Sensing of Environment*. Elsevier, 114(2), pp. 449–465. doi: 10.1016/J.RSE.2009.10.002.
- Wan, Z. (2013) *Collection-6 MODIS Land Surface Temperature Products Users' Guide*. Available at: https://icdc.cen.uni-hamburg.de/fileadmin/user_upload/icdc_Dokumente/MODIS/MODIS_Collection6_LST_products_UsersGuide.pdf (Accessed: 31 July 2018).
- Wang, F. (1994) *The use of artificial neural networks in a geographical information system for agricultural land-suitability assessment*, *Environment and Planning A*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.457.8304&rep=rep1&type=pdf> (Accessed: 28 August 2018).