

Pronóstico de temperatura mínima del aire en zonas susceptibles a heladas radiativas en Colombia.

Oscar David Rivera Ricaurte ^a

^a Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. odriverar@unal.edu.co

Resumen

Se muestra el diseño de un modelo de aprendizaje automático que usa datos espaciotemporales (ERA5 en niveles individuales) de variables meteorológicas asociadas a la variación de la temperatura mínima del aire para predecirla en zonas susceptibles a heladas radiativas en Colombia. Para esto se hace una selección de variables, seguida de unas pruebas para definir el mejor algoritmo de modelación y los horizontes de pronóstico más adecuados. Los modelos seleccionados son optimizados en sus hiper parámetros y finalmente se entrenan y validan. Los resultados de las simulaciones con un modelo lineal multivariado con regularización tipo ridge y un multi layer perceptron muestran resultados destacables, buen aprendizaje y poco sobre ajuste.

Palabras clave: pronóstico; heladas radiativas; temperatura mínima; modelos de aprendizaje automático; regresión lineal multivariada; multi layer perceptron.

1 Introducción

La temperatura mínima del aire es una variable meteorológica importante en distintos procesos productivos, especialmente los sectores relacionados con la crianza de animales y la siembra de alimentos. Esto se debe a que tanto animales como plantas tienen cierta tolerancia a las condiciones ambientales extremas y por ejemplo, descensos anómalos en la temperatura del aire pueden llevar a disminuir el rendimiento que se conseguiría bajo condiciones medias [1].

Esto es especialmente notable en el territorio colombiano cuando se habla de heladas, fenómenos meteorológicos que se caracterizan por magnitudes de temperatura mínima del aire iguales o inferiores a 0 °C. Durante estas se pueden notar capas de congelamiento sobre (y en algunos casos al interior de) sembradíos y pastos que interfieren con el metabolismo de estos organismos y en algunos casos llegan a causar la muerte de sus tejidos [1], [2]. Existen estrategias que ayudan a mitigar o evitar los daños causados por el congelamiento de las plantas, pero estos dependen del conocimiento que se tenga sobre el fenómeno en cuanto a las zonas y periodos con mayor frecuencia [3]. Ante esta necesidad, el diseño de modelos de pronóstico que consideren la espacialidad de la temperatura mínima surgen como un insumo de gran utilidad.

En un principio este trabajo plantea el desarrollo de un modelo de pronóstico de heladas el cual podría ayudar a mejorar los tiempos de reacción ante posibles amenazas de ocurrencia de estos eventos. Sin embargo, dado que se desea tener una cobertura espacial en todo Colombia, que los datos de estaciones meteorológicas de IDEAM no poseen registros con la calidad deseada y que estos fenómenos

(hasta el momento) no suelen ser representados con fiabilidad por datos espaciales debido a su extensión (de unos pocos kilómetros según [4]), el objetivo se modifica y centra en construir un modelo para la predicción de la temperatura mínima del aire en zonas susceptibles a heladas en Colombia. Se tienen como metas de modelación una precisión de alrededor del 80% medido en porcentaje medio absoluto de error (MAPE), y con un horizonte de pronóstico alrededor o por encima de 6 horas.

2 Datos y Metodología

Para la selección de los datos se tienen en cuenta características que permitan que a futuro el modelo diseñado pueda evolucionar y pronosticar heladas en la zona de estudio. Primero se recopila información de los tipos de heladas que ocurren en Colombia y sus determinantes para posteriormente establecer aquellos requisitos a cumplir para seleccionar el conjunto de datos.

En Colombia suceden casi exclusivamente heladas de tipo radiativo, las cuales son favorecidas por condiciones ambientales de tipo geomorfológico, edafológico, ecológico y meteorológico [2], siendo estas últimas las de mayor variación en el corto plazo y las que se consideran determinantes en la ocurrencia de estos fenómenos, por lo tanto, se buscarán predictores de este tipo.

Además, para garantizar que se está pronosticando la temperatura en zonas susceptibles a heladas radiativas se usaran datos en alturas superiores a 2400 m.s.n.m., dado que según [2] este fenómeno solo se presenta a estos niveles. Esta área se muestra en la Fig.1, donde los pixeles rojos representan el área a estudiar.

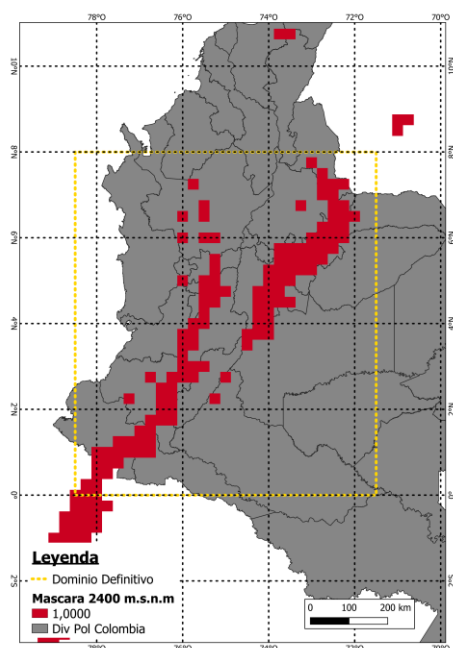


Figura 1. Área de estudio.

Fuente: Elaboración propia.

2.1 Datos

Para seleccionar la base de datos se tienen en cuenta distintos requisitos:

- En primer lugar, que los datos sean espaciotemporales y tengan una cobertura en toda el área de estudio.
- Se busca que la resolución espacial sea lo más alta posible.
- Debe tener una frecuencia de muestreo horaria puesto que las temperaturas mínimas más extremas se dan en horas previas al amanecer y por procesos que favorecen el enfriamiento nocturno. También un periodo de muestreo lo más prolongado posible.
- Se deben tener registros de distintas variables meteorológicas, al menos: temperatura mínima del aire, temperatura del punto de rocío, velocidad del viento, radiación y humedad relativa o similares, pues estas han mostrado tener relación con los descensos extremos de temperatura mínima [2], [4]–[6].

Con estas pautas se selecciona el reanálisis climático ERA5 en su producto single levels, que mezcla datos de simulaciones y observaciones y tiene una resolución espacial de 0.25° (~ 28 Km), frecuencia horaria, cobertura global y en periodo desde 1979 hasta la actualidad.

2.2 Metodología

2.2.1 Selección de variables

En primera instancia se revisa literatura para determinar aquellas variables con más relación con el predictando [2],

[4]–[6], las cuales se encuentran en la Tabla 1 con sus respectivas abreviaciones y unidades.

Tabla 1
Variables objeto de selección.

Variable	Abreviación	Unidades
Altura de la capa límite	blh	m
Cobertura de nubes a baja altura	lcc	%
Presión atmosférica	sp	Pa
Precipitación total	tp	M
Radiación térmica del suelo	str	$J\ m^{-2}$
Temperatura mínima del aire	mn2t	K
Temperatura del punto de rocío	d2m	K
Vapor total en la columna de aire	tcwp	$kg\ m^{-2}$
Viento meridional	u10	$m\ s^{-1}$
Viento zonal	v10	$m\ s^{-1}$

Fuente: Elaboración propia

A este conjunto de variables se les aplican principalmente dos pruebas para determinar cuales tienen mayor correlación con la temperatura mínima del aire y son lo mas independientes posible. La primera es una matriz de correlación la cual no fue concluyente para todos los parámetros, por lo cual se revisan los histogramas de los datos estandarizados y se comparan cada uno con el de la temperatura mínima para descartar aquellos índices con menor similitud en su distribución. También se aplica el método no supervisado de análisis de componentes principales el cual permitiría descartar mas variables y disminuir la cantidad de predictores, no obstante, esta prueba se usó para analizar el problema desde un punto de vista más físico.

2.2.2 Pre selección del modelo y horizonte de pronóstico.

Son distintos los modelos de los que se dispone en la librería scikit-learn para la solución de problemas tipo regresión, por lo tanto, se considera evaluar la gran mayoría de estos y seleccionar el o los de mejor desempeño.

Para estas pruebas se selecciona un periodo acotado de datos (solo el año 2020), puesto que se ajustarán varios algoritmos y esto permite reducir el tiempo y recursos de computación necesarios.

Los distintos modelos también serán evaluados en su desempeño ante diferentes horizontes de pronóstico de 1,2,3,6,9,12,16,20 y 24 horas.

Es importante resaltar que en esta sección no se usa como variable predictora la temperatura mínima del aire, dado que esta mejora en gran medida al desempeño de los modelos y no permite notar las diferencias entre ellos fácilmente.

Los algoritmos por comparar son regresión lineal con regularización ridge (LMR), k vecinos mas cercanos (KNN), arboles de decisión (CLF), random forest (RF), bagging (BAG) y multi layer perceptrón (MLP), los cuales son instanciados sin cambiar sus hiperparámetros. También se hicieron pruebas con support vector machine, pero se

observó una demanda de recursos muy elevada a comparación de los otros modelos y se decide descartarlo. La división de las series de datos en entrenamiento y validación es de 0.67 y 0.33 del total de los datos respectivamente y los modelos se ajustan y validan para todos los pixeles con las métricas de coeficiente de determinación en entrenamiento y validación, raíz del error cuadrático medio en validación y tiempo de cómputo.

2.2.3 Optimización de hiperparámetros

Al igual que en preselección, se usa únicamente el año 2020 para reducir la demanda de recursos computacionales.

La librería scikit-learn cuenta con la herramienta GridSearchCV que permite evaluar distintas combinaciones de hiper parámetros de los respectivos algoritmos y encontrar aquellas que mejoran el desempeño de estos.

Teniendo en cuenta que GridSearchCV realiza una cantidad alta de iteraciones, en este caso no se implementa en cada uno de los pixeles de la zona de estudio, sino que a partir de la información obtenida en los cálculos de preselección se usa la serie de datos correspondiente a la latitud y longitud donde los modelos presentaron el peor desempeño para el ajuste de los hiper parámetros.

Al igual que en preselección se usan las métricas coeficiente de determinación en entrenamiento y validación, raíz del error cuadrático medio en validación y tiempo de cómputo para comparar el desempeño del modelo sin optimizar sus hiperparámetros y el optimizado.

2.2.4. Entrenamiento y validación

Para esta etapa se usa el conjunto completo de datos. Sin embargo, para reducir el tiempo de ajuste se realiza una integración de todas las variables en el espacio por su promedio con el fin de ajustar los modelos una única vez y que esta solución sea aplicada a todos los pixeles de la región a simular.

El tamaño de la validación es igualmente de 0.33 y las métricas a evaluar son los coeficientes de determinación en entrenamiento y validación y el error porcentual medio absoluto.

3 Resultados

3.1 Selección de variables

La primera prueba aplicada es una matriz de correlación a partir de la cual se revisan aquellas variables que tengan un coeficiente por encima de 0.5 entre sí y con la menor relación con la temperatura mínima del aire. Esta matriz se muestra en la Fig. 2 donde los colores más cálidos representan una conexión directamente proporcional entre los parámetros, los más fríos una relación inversa y los colores más opacos indican desconexión.

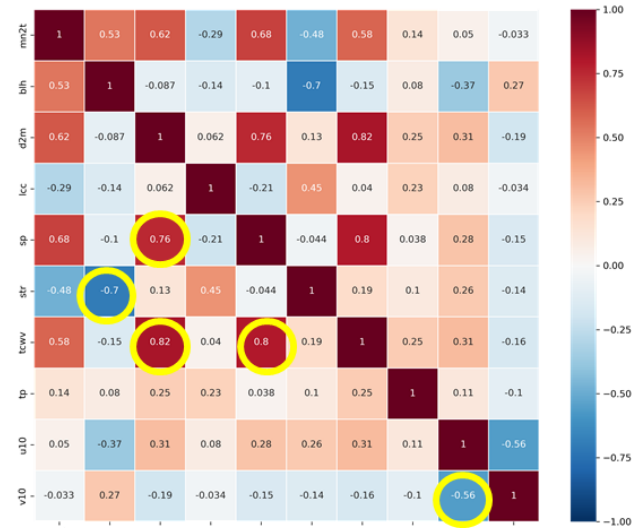


Figura 2. Matriz de correlación de variables meteorológicas
Fuente: Elaboración propia

Se observa que las variables sp y tcwv tienen alta correlación con d2m y teniendo en cuenta que no se relacionan tan bien con la mn2t, estas dos son descartadas y se conserva d2m. Otras variables con alta linealidad son u10-v10 y str-blh, no obstante, las correlación de estas con mn2t no tiene diferencias tan notables y por esto se revisan los histogramas marginales de estas variables y mn2t.

La Fig. 3 muestra los histogramas marginales de las variables estandarizadas, donde las líneas azules representan la mn2t y las naranjas la variable meteorológica a comparar.

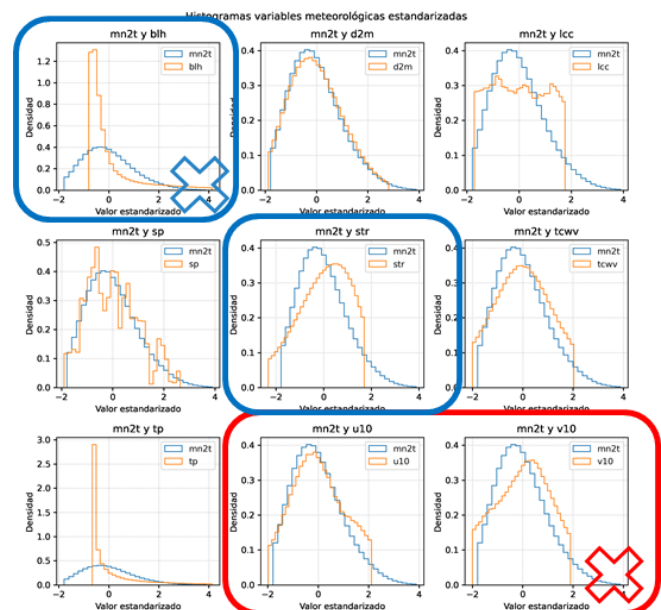


Figura 3. Histogramas marginales de las variables meteorológicas.
Fuente: Elaboración propia.

En estos gráficos se hace énfasis en la revisión de blh, str u10 y v10, apreciando que entre blh y str esta última

variable tiene un histograma con una distribución notablemente más cercana a la *mn2t* con lo cual se descarta a *blh* como predictor. Entre los histogramas de *u10* y *v10* las diferencias son menos abruptas, pero finalmente se descarta *v10*, lo cual tiene sentido teniendo en cuenta que la meteorología de Colombia depende en mayor medida de vientos meridionales (*u10*) dada la cercanía al Ecuador, al cinturón de vientos alisios y la zona de convergencia intertropical.

Finalmente, la Fig. 4 muestra el análisis de componentes principales de las variables meteorológicas. Cabe aclarar que esta es una integración con el comportamiento promedio de los PCAs en la zona de estudio, dado que estos fueron ajustados en cada píxel.

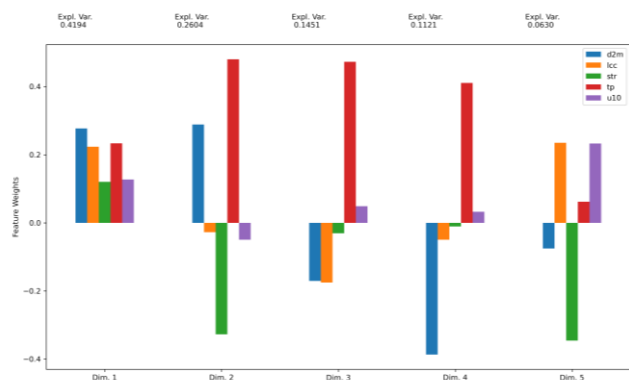


Figura 4. Varianza explicada y peso de las variables meteorológicas de los PCAs
Fuente: Elaboración propia.

En esta se observa que la mayoría de la varianza es explicada por los 3 primeros PCA y que, en estos, las variables con mayor peso son *d2m*, *tp* y *str* y quedan en un segundo plano *lcc* y *u10*. Si analizamos esto desde el punto de vista físico esto indicaría que las variaciones más importantes en el ambiente atmosférico en estas zonas son aquellas que tienen que ver con la humedad presente, representada por *d2m* y *tp* y también por la cantidad de radiación de onda larga o en otros términos del calor liberado por el suelo. Como se menciona en la sección de metodología este método se evalúa más con la intención de entender mejor que variables pueden ser de mayor importancia física para el fenómeno, pero si se deseara acotar el modelo se podrían seleccionar estas tres variables o los primeros 3 PCA como se recomienda en [7].

3.2 Preselección del modelo y horizonte de pronóstico

Las pruebas conducidas durante esta etapa se pueden ver resumidas en la Fig. 5 donde se muestra una integración hecha en el espacio a partir de la mediana de las distintas métricas evaluadas en cada modelo.

En primer lugar, se puede observar que el modelo que presenta las peores métricas en cuanto a RMSE y la comparación de sus coeficientes de determinación en

entrenamiento y validación es el modelo de árboles de decisión con lo cual es el primer modelo descartado.

Por otra parte, los algoritmos de random forest y bagging si bien no muestran un sobre ajuste tan marcado como árboles de decisión, si se puede ver que en comparación a los otros tres modelos estos pueden presentar más problemas de este tipo y por lo tanto se descartan también.

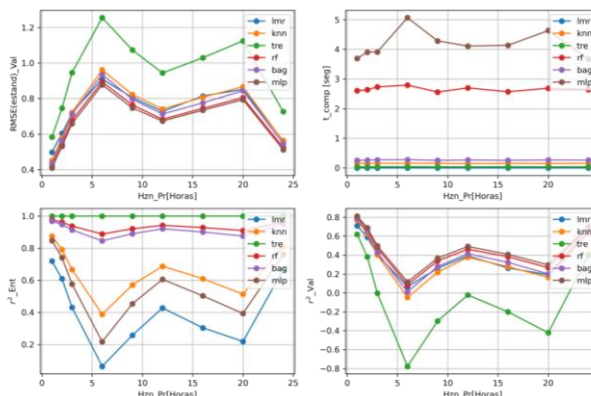


Figura 5. Comparativa de los modelos en la preselección
Fuente: Elaboración propia.

Finalmente se tienen tres algoritmos como candidatos para la modelación y teniendo en cuenta las ventajas que puede tener LMR en cuanto a interpretabilidad de los resultados y simplicidad de la solución y MLP en cuanto a capacidad de aprendizaje y simulación de no linealidades, se conservan estos dos modelos para realizar el ajuste de hiperparámetros y la modelación.

En cuanto al horizonte de pronóstico, los resultados permiten concluir que los mejores resultados se alcanzan rezagando las variables por 1 hora, pero, teniendo en cuenta que a futuro se espera convertir este modelo en un predictor de heladas, se considera tomar el segundo periodo con mejores resultados que es el de 24 horas, dado que permite tener un tiempo de reacción más prolongado ante eventos extremos de temperatura.

Este resultado se puede explicar en cuanto a que la temperatura y las demás variables meteorológicas tienen un ciclo diurno relativamente estable a largo plazo que depende de la cantidad de radiación solar, por lo tanto, tendrán más similitud aquellos datos con una separación de 24 horas que con una de 12 o 6 horas dado que las condiciones de irradiación serán más similares cuando la separación de los registros es de 1 día y para el modelo será más fácil asimilar la relación entre las variables y el pronóstico con este periodo de rezago.

3.3 Optimización de hiperparámetros

Como se menciona anteriormente los modelos a evaluar en esta sección serán LMR y MLP. Para las iteraciones con la herramienta GridSearch para el LMR se seleccionan como hiperparámetros a optimizar el ajuste del intercepto, la magnitud de Alpha y el método de solución para los coeficientes. En el caso de MLP se tienen la función de activación, el método de solución, la tasa de aprendizaje, si

hay calentamiento para el modelo y la arquitectura de este. En la Tabla 2 se muestran los valores de los hiperparámetros que fueron objeto de iteración y también aquellos que GridSearch entrega como óptimos.

Es importante destacar que en el caso de MLP primero se hicieron pruebas con todos los parámetros a excepción de la arquitectura y posteriormente se evaluó esta por separado, además que para este algoritmo se activó la función de *early_stopping* que permite detener las iteraciones cuando la solución no está mejorando sustancialmente.

Tabla 2.
Hiperparámetros objeto de optimización

Algoritmo	Hiperparámetro	Valores	Selección
LMR	Intercepto	Verdadero, Falso	Falso
	Alpha	0.001, 0.01, 0.1, 1, 5, 10	10
	Solución	'svd', 'cholesky', 'lsqr', 'sparse_cg', 'sag', 'saga'	Sag
MLP	Activación	'logistic', 'tanh', 'relu'	Logistic
	Solución Aprendizaje	'lbfgs', 'sgd', 'adam', 'invscaling', 'adaptive'	Lbfgs
	Calentamiento	Verdadero, Falso	Falso
	Arquitectura	(6,6), (6,12), (6,3), (3,6,3), (3,3), (12,6), (3,3,3), (6,6,6)	(12, 6)

Fuente: Elaboración propia.

Se resalta que en LMR se selecciona un intercepto nulo, con lo cual se puede suponer que hay pocas variables exógenas que estén siendo ignoradas en este modelo y la mayoría de la variabilidad del predictando es captada por los predictores. También se destaca que el solucionador para MLP no utiliza una tasa de aprendizaje y por lo tanto este hiperparámetro es ignorado.

Para evaluar que tan bueno es el desempeño del modelo optimizado y sin optimizar se usan las mismas métricas que en la sección de preselección del modelo, las cuales se encuentran resumidas en la Tabla 3.

Tabla 3.
Métricas para los modelos con hiperparámetros optimizados y sin optimizar.

Métrica	LMR	LMR-OP	MLP	MLP-OP
R2 Entrenamiento	0.890	0.890	0.901	0.901
R2 Validación	0.856	0.856	0.865	0.865
RMSE	0.352	0.351	0.341	0.341
t computo [s]	0.002	0.016	0.009	0.003

Fuente: Elaboración propia.

Esta comparativa permite observar que los modelos tienen pocas diferencias entre sí, lo cual se puede deber a que se tiene una buena extensión en las series de tiempo para su entrenamiento y los hiperparámetros no influyen tanto en la solución final. En cuanto a LMR se decide usar el modelo sin optimizar y solo seleccionando el intercepto como nulo dado que la solución con los parámetros de GridSearch es ligeramente menos exacta en cuanto a RMSE y necesita más tiempo de cómputo. Por el contrario, en el MLP la herramienta logra escoger hiperparámetros que, si bien no mejoran el desempeño en error, si lo hacen reduciendo en tiempo de cómputo, probablemente debido a que el modelo converge más rápidamente a una solución y necesita hacer menos iteraciones.

En resumen, los modelos e hiperparámetros que serán usados en la etapa de modelación son LMR con intercepto falso y MLP con dos capas de 12 y 6 neuronas respectivamente, solucionador tipo lbfgs, sin calentamiento y con función de activación logística.

3.4 Modelación.

Como se menciona en la sección 2.2.4, para el entrenamiento de los modelos LMR y MLP se usa una serie de datos que es una integración espacial que se hace promediando los datos en latitud y longitud con lo cual solo se conserva la dimensión del tiempo. Los datos de entrenamiento son rezagados 24 horas siguiendo los resultados obtenidos en secciones anteriores. En total se tienen 354240 registros horarios en todas las variables que son separados y se conserva un 0.33 de su total para la validación y un 0.67 para entrenamiento, es decir 237340 registros son usados para esta fase.

Al igual que en pasos previos los datos son estandarizados con la herramienta *standar scaler* de *scikit learn* para evitar que la magnitud de las variables intervenga en los coeficientes asignados.

La Tabla 4 hace un resumen de los coeficientes ajustados por LRM para cada una de las variables predictoras.

Tabla 4
Coeficientes de LMR para el entrenamiento del modelo.

Variable	Coeficiente
d2m	0.023
lcc	-0.020
mn2t	0.906
str	-0.043
tp	0.040
u10	-0.013

Fuente: Elaboración Propia

Estos coeficientes permiten obtener conclusiones similares a las del análisis de componentes principales, es decir que las variables que más predominan en la varianza del ambiente atmosférico y en este caso de la temperatura mínima del aire son la tp, la d2m y la str. Además, el signo de los coeficientes permite suponer la reciprocidad de la relación entre las respectivas variables y la mn2t. Se observa que

mayores niveles de t_p y d_2m pueden estar asociados a temperaturas más cálidas lo cual se puede deber a que una mayor cantidad de humedad en la atmósfera hace que su capacidad calorífica aumente y conserve mejor la energía térmica, así como mayores niveles de precipitación se deben a procesos de convección más profunda y por lo tanto de condiciones en las cuales los niveles de aire cercanos a la superficie tienen una alta temperatura. En el caso de la str el comportamiento se explica en cuanto a que una mayor liberación de calor del suelo al aire implica una pérdida acelerada de la energía que entrega el sol durante el día y por lo tanto el enfriamiento puede ser mayor.

Una vez se entrena el modelo y se ajusta a los datos de validación se observa que los dos algoritmos entregan respuestas idénticas. Por tal motivo a continuación se muestran solo los resultados obtenidos con MLP, los cuales se resumen en los mapas de la Fig. 6, donde el panel a representa la diferencia entre los coeficientes de determinación en entrenamiento menos el de validación, el b muestra las magnitudes del r^2 en validación y el c representa el error porcentual absoluto medio en la etapa de validación.

La Fig. 6 a permite observar que en general los modelos tienen valores cercanos en sus etapas de ajuste y validación con diferencias máximas de 0.035, además teniendo en cuenta que el modelo es entrenado con solo una serie de tiempo que integra todo el espacio, se concluye que el modelo no tiene problemas de sobre ajuste y puede ser aplicado en toda la región de estudio.

En cuanto al coeficiente de determinación en validación (Fig. 6 b) se muestra en general un buen ajuste del modelo con valores mínimos de alrededor de 0.8 y alcanzando niveles incluso de 0.95.

El error porcentual también está en valores aceptables, siendo de alrededor de 2 a 6 por ciento, con solo un píxel que tiene un valor anómalo en comparación a los demás.

4 Discusión y Conclusiones

Los resultados obtenidos tanto para LMR como para MLP muestran que se han alcanzado modelos robustos, con poco o nulo sobre ajuste, que son capaces de hacer pronósticos fiables de temperatura mínima del aire con un horizonte de 24 horas. Las métricas evaluadas permiten concluir que los objetivos planteados en este trabajo fueron cumplidos y que permiten pensar en avanzar hacia el desarrollo de un modelo de pronóstico de heladas radiativa en Colombia.

Los coeficientes de la LMR destacan como mejor predictor a la propia mn_{2t} rezagada 24 horas y se podría pensar que la fase inicial de selección de variables fue desarrollada en vano, no obstante esa sección en conjunto con la información de la LMR permiten elaborar conjeturas sobre la relación entre la mn_{2t} y las variables meteorológicas que modifican su comportamiento, lo cual deja abierta la puerta a nuevas investigaciones que busquen no solo correlaciones entre estos parámetros sino también conexiones causales.

A futuro este modelo también puede ser objeto de experimentación, por ejemplo, modificando el horizonte de

pronóstico para observar con cuantos días de antelación este puede tener una exactitud deseada.

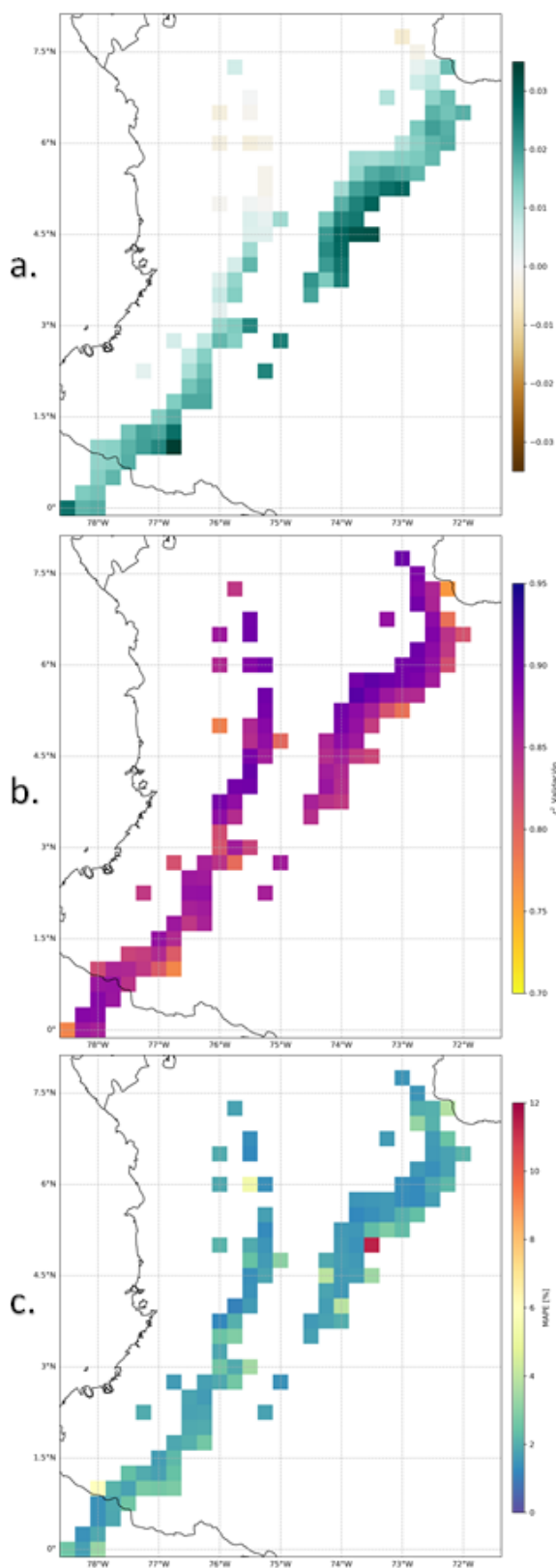


Figura 6. Métricas del modelo MLP para la región modelada. a. Diferencia en el r^2 para entrenamiento y validación, b. r^2 para la validación y c. MAPE validación
Fuente: Elaboración Propia

Teniendo en cuenta que el horizonte con mayor precisión es el de 1 hora, se puede explorar la posibilidad de hacer simulaciones con este periodo y que las salidas sean tomadas por el modelo para extender los resultados pronosticados al tiempo que se desee, evaluar hasta cuando es efectivo este método y si mejora las simulaciones hechas con el horizonte de 24 horas. Para esto se puede explorar la posibilidad de usar también el algoritmo de redes neuronales tipo LSTM que se enfoca en problemas donde las respuestas dependen de los antecedentes.

Otro experimento sugerido es el uso de distintos conjuntos de predictores para revisar cómo se comporta la respuesta del modelo, similar a un análisis de sensibilidad, enfocándose en que parámetros mejoran considerablemente los resultados del modelo en términos de precisión de los valores entregados y de que tanto ayudan a prolongar el horizonte de pronóstico.

Dado que este es un problema de variables meteorológicas las cuales están expuestas a cambios relacionados al cambio climático y el calentamiento global, se pueden tomar distintos periodos de tiempo para entrenar el modelo y observar si este es susceptible a este tipo de alteraciones o por el contrario es un algoritmo que puede ser usado con confianza por un tiempo prolongado.

Una conclusión importante es que, aunque en un principio se puede pensar que el LMR puede ser recomendable por encima de MLP por tiempo de cómputo, las pruebas realizadas muestran que, aunque pueda sonar poco lógico MLP termina ajustándose más rápido a los datos y reduce los recursos necesarios para su entrenamiento. Además, cuando observamos las Fig. 7, la cual en su panel a muestra la curva de aprendizaje con el 10% de los datos para LMR y en el b la misma curva para MLP, se puede concluir que MLP converge a una solución estable, con mayor rapidez y más precisa que el LMR. Esta figura también permite observar que para el entrenamiento de MLP son necesarios menos datos y por lo tanto se puede entrenar con un periodo reducido o para zonas donde la cantidad de datos sea menor. En conclusión, para este problema es más recomendable el uso de MLP.

La calidad de los resultados también puede deberse a la fiabilidad del conjunto de datos usado, con lo cual se pueden explorar sus bondades en otros campos de las geociencias dado que este reanálisis cuenta con una gran cantidad de variables de distinto tipo.

Dado que a futuro se espera modificar este modelo para el pronóstico de heladas radiativas en Colombia también es importante revisar no solo las métricas medias sino también el comportamiento de las series de tiempo. Para esto se hace la Fig. 8, a partir de una integración en el espacio con el promedio de los datos pronosticados y los reales en el dominio seleccionado. Se puede observar que el modelo representa bien el comportamiento medio de la temperatura

real, sin embargo, cuando observamos los extremos tanto de temperatura máxima como de temperatura mínima, se observa que hay secciones donde el modelo tiene fallas más notables.

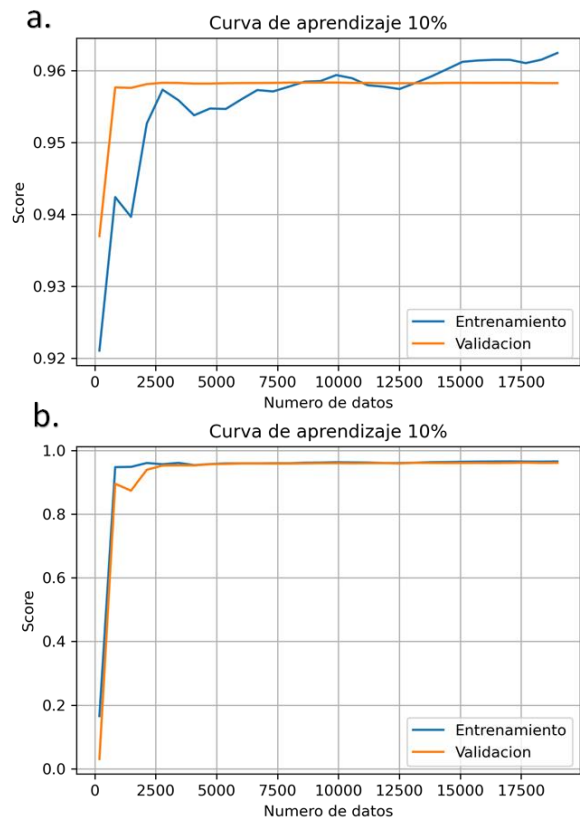


Figura 7. Curvas de aprendizaje con el 10% de los datos para a. LMR y b. MLP.

Fuente: Elaboración propia

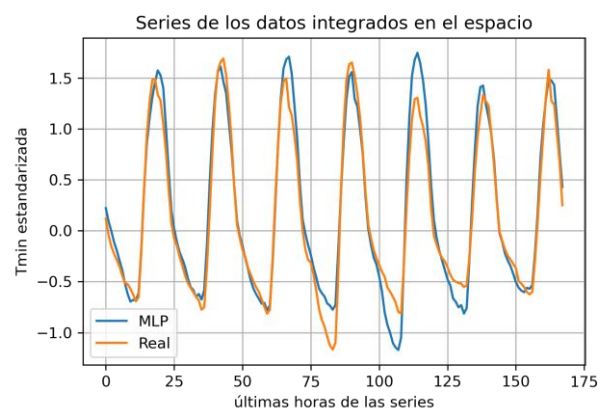


Figura 8. Serie de tiempo de las últimas horas pronosticadas por MLP.

Fuente: Elaboración Propia

Teniendo en cuenta que las heladas radiativas son fenómenos de temperaturas mínimas extremas se puede pensar que por el momento este modelo no sería capaz de representar muy bien estos eventos y para su evolución,

debe ser entrenado teniendo un mejor cuidado de como se entrena, con el fin de que calcule de mejor manera que conjunto de variables indican que se debe entregar una respuesta extrema. Para esto a futuro se investigará el uso de datos de temperatura mínima in situ del IDEAM o de otro tipo de datos satelitales como MODIS y su temperatura en superficie la cual tiene una mejor resolución y puede ayudar a diagnosticar los puntos y épocas en los cuales hay heladas radiativas y tener en cuenta esta información para el desarrollo y entrenamiento del modelo.

Otra pregunta abierta que debe mejorarse a futuro es explicar la variación en el espacio de la respuesta de la mn2t simulada y entender de mejor manera las variables meteorológicas o biofísicas que explican este comportamiento.

5 Referencias

- [1] FAO, “El daño producido por las heladas: Fisiología y Temperaturas críticas,” *Protección contra las heladas Fundam. práctica y Econ.*, vol. 15, no. 4, pp. 73–98, 2010, [Online]. Available: <http://www.fao.org/docrep/012/y7223s/y7223s00.htm>.
- [2] IDEAM, “Actualización Nota Técnica Heladas 2012,” *Ideam*, p. 79, 2012, [Online]. Available: <http://www.ideam.gov.co/documents/21021/21147/Documento+FINAL+actualizacion+nota+tecnica+heladas.pdf/e10a0183-62e6-410a-8e96-7e0739f6f06b>.
- [3] FAO, *Protección contra las heladas : fundamentos , práctica y economía*, 1st ed. 2010.
- [4] INIA, “Heladas. Factores, tendencias y efectos en frutales y vides,” p. 102, 2020.
- [5] E. Torres, “Caracterización de las temperaturas extremas del aire para el cultivo de papa en la Sabana de Bogotá,” *Repositorio UNAL*, p. 296, 2020, [Online]. Available: <https://repositorio.unal.edu.co/handle/unal/78279>.
- [6] M. Saavedra and K. Takahashi, “Physical controls on frost events in the central Andes of Peru using in situ observations and energy flux models,” *Agric. For. Meteorol.*, vol. 239, pp. 58–70, 2017, doi: 10.1016/j.agrformet.2017.02.019.
- [7] T. Doan and J. Kalita, “Selecting Machine Learning Algorithms using Regression Models,” in *International Conference on Data Mining Workshop (ICDMW)*, 2016, no. November, pp. 1498–1505, doi: 10.1109/ICDMW.2015.43.