

PREDICCIÓN PROBABILÍSTICA DE LA PRECIPITACIÓN MENSUAL EN LA CUENCA DEL RÍO BERMEJO

Sabrina Ayala^{1,2,3}, Marcela González^{2,3}
sabrina.ayala@cima.fcen.uba.ar.

¹Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos. Buenos Aires, Argentina.

²CONICET – Universidad de Buenos Aires. Centro de Investigaciones del Mar y la Atmósfera. Buenos Aires, Argentina.

³CNRS – IRD – CONICET – UBA. Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI). Buenos Aires, Argentina.

Palabras clave: precipitación, Bermejo, aprendizaje estadístico, pronóstico probabilístico

1) INTRODUCCIÓN

El sector argentino de la Cuenca del Río Bermejo (CRB) se ubica en sector norte del país y comprende las subcuencas Bermejo Superior y San Francisco (BS y SF, respectivamente, ubicadas en el extremo noroeste) y la subcuenca Media-Inferior (BMI, que se extiende hacia el noreste del país). Se trata de una región con un régimen hidrológico pluvial principalmente concentrado en la estación cálida austral. Además, es una zona relevante en la producción agropecuaria de secano de sorgo, soja, algodón, maíz y porotos (Subsecretaría de Agricultura, 2022).

El objetivo de este estudio es diseñar un conjunto de modelos estadísticos de predicción de la precipitación mensual para la CRB. Para ello, se considerarán un conjunto de técnicas lineales y no lineales de aprendizaje estadístico, lo que permitirá elaborar un pronóstico probabilístico de la precipitación mensual a mediano plazo durante el semestre de mayores acumulados pluviométricos (octubre-marzo).

2) DATOS Y METODOLOGÍA

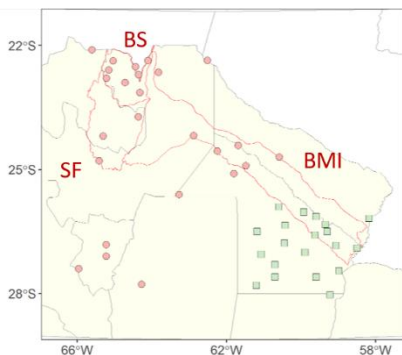


Figura 1: Localización de las estaciones que conforman los clusters 1 (rosa) y 2 (verde). En rojo se delimitan las subcuencas BS, SF, y BMI de la CRB.

Se emplearon datos de precipitación mensual provenientes del Servicio Meteorológico Nacional (SMN), el Sistema Nacional de Información Hídrica (SNHI), y la Administración Provincial del Agua de Chaco (APA), correspondientes al semestre cálido entre octubre y marzo del período 1984/1985-2019/2020. Los datos atravesaron un proceso de control de calidad, admitiendo solamente aquellas estaciones de medición con una cota superior de faltantes del 10%. Las estaciones fueron agrupadas en dos áreas climáticamente homogéneas siguiendo la metodología de k -medoides, empleando como valor óptimo $k = 2$ considerando los criterios de suma total de cuadrados intra-cluster y ancho de silueta (Figura 1). En base a los clusters hallados, se obtuvieron las series temporales medias areales de precipitación mensual para el semestre cálido.

Se exploraron diversos predictores de la precipitación para cada mes del semestre y para ambos clusters. Esto se realizó mediante la construcción de índices que representaran forzantes

climáticos de gran escala y escala regional, y que exhibieran una correlación lineal significativa al 95% de confianza (test T-student) con la precipitación mensual del mes inmediatamente posterior. Los índices se construyeron a partir de datos medios mensuales del reanálisis ERA5 (Hersbach et al., 2019) de temperatura superficial del mar, altura geopotencial (en 1000 hPa, 700 hPa, y 500 hPa), agua precipitable, flujo de radiación de onda larga saliente (OLR), y componentes zonal y meridional del viento en 850 hPa y 200 hPa. Los predictores fueron seleccionados manteniendo la independencia lineal entre ellos, y empleando la metodología de regularización de contracción mínima absoluta (LASSO).

En base a los predictores definidos para cada cluster, se entrenaron los modelos estadísticos de predicción de la precipitación mensual mediante la estrategia de validación cruzada de un elemento sobre el período 1984/1985-2019/2020. Las técnicas empleadas fueron la Regresión Lineal Múltiple (RLM), Modelos Aditivos Generalizados (GAM), Regresión de Vectores de Soporte (SVR), y Redes Neuronales Artificiales multicapa *feedforward* (ANN). Además, se seleccionaron aquellos modelos que presentaran un coeficiente de determinación ajustado ($AdjR^2$) mayor al 40% sobre la muestra de entrenamiento. El desempeño de los modelos sobre el pronóstico determinístico de la precipitación mensual para cada cluster fue verificado mediante la utilización de diagramas de Taylor. Finalmente, se evaluó la capacidad predictiva del pronóstico probabilístico de la precipitación por terciles (inferior a lo normal “*sub*”, normal “*nor*”, y superior a lo normal “*sobre*”) empleando como métricas de evaluación al Heidke Skill Score (HSS), diagramas de tendencia (DT), Brier Skill Score (BSS), y el área bajo la curva Relative Operating Characteristic (AUC). Dichas métricas sintetizan el desempeño de los pronósticos probabilísticos en términos de proporción de aciertos, resolución, sesgo no condicionado, confiabilidad y discriminación (Mason y Stephenson, 2008).

3) RESULTADOS

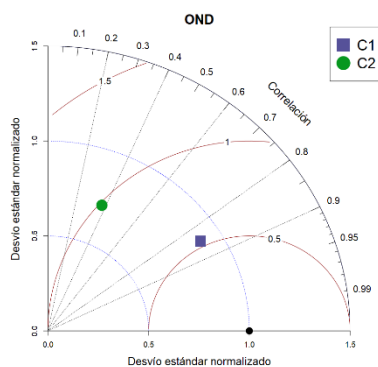


Figura 2. Diagrama de Taylor de los ensambles multimodelo del cluster 1 (cuadrado violeta) y 2 (círculo verde) en los meses que componen la primavera (OND).

Utilizando los datos del reanálisis ERA5, se definieron los predictores de la precipitación mensual para ambos clusters de la región de estudio mediante el análisis de correlación en conjunto con la técnica LASSO. Los mismos corresponden a cuatro grupos generales: en primer lugar, se hallaron predictores que representan la variabilidad en latitudes bajas de los océanos Pacífico e Índico. En segundo lugar, se identificaron predictores que representan la acción de trenes de onda en latitudes extratropicales. En tercer lugar, se identificaron predictores que representan las variaciones en intensidad y posición del anticiclón semipermanente del océano Atlántico sur. Por último, se definieron predictores vinculados a la evolución estival del sistema monzónico sudamericano y su aporte de humedad hacia el norte argentino.

Entre los modelos entrenados considerando los predictores identificados, se encontró que los modelos de tipo SVR y ANN muestran predominantemente valores de $AdjR^2$ mayores al 40% especialmente para el cluster 1, mientras que los modelos RLM y GAM muestran valores más bajos para esta métrica (Figura no mostrada). Si bien todos los modelos tienden a subestimar la variabilidad observada (desvío estándar de los modelos menor al observado), los modelos de ANN tienden a reproducir más fielmente la variabilidad de la precipitación mensual. En particular, el ensamble multimodelo para los meses de primavera (octubre a diciembre, OND) muestra valores más favorables para el cluster 1 que para el cluster 2 en términos de desvío estándar normalizado, correlación con los valores observados, y error cuadrático medio normalizado (Figura 2). Para el verano (enero a marzo, EFM), el rendimiento de los modelos

del cluster 1 disminuye, y éstos se acercan al desempeño de los modelos para el cluster 2 (Figura no mostrada).

CLUSTER 1										
	HSS	Sub			Nor			Sobre		
		DT	BSS	AUC	DT	BSS	AUC	DT	BSS	AUC
Oct	0.34		*	*			*	*	*	*
Nov	0.50	*	*	*			*	*	*	*
Dic	0.11		*	*			*	*	*	*
Ene	0.15			*				*		*
Feb	0.42	*	*	*	*	*	*	*	*	*
Mar	0.15			*			*	*	*	*

CLUSTER 2										
	HSS	Sub			Nor			Sobre		
		DT	BSS	AUC	DT	BSS	AUC	DT	BSS	AUC
Oct	0.42	*	*	*		*	*	*	*	*
Nov	0.23	*		*			*			*
Dic	0.07						*	*		*
Ene	0.0			*						*
Feb	0.27	*	*	*		*	*	*	*	*
Mar	0.07			*			*			*

Tabla I. Síntesis de las métricas de evaluación de desempeño para el pronóstico probabilístico de ambos clusters y para cada mes del semestre cálido. Las celdas con asterisco indican resultados favorables en comparación a la climatología para DT, BSS, y AUC.

En cuanto a la verificación desde el enfoque probabilístico, se emplearon diversas métricas incluidas en la Tabla I. En la misma, se indica para cada cluster los meses en los cuales el pronóstico probabilístico mostró resultados favorables comparado a la climatología (BSS > 0, AUC > 0.5, sesgo en DT < 10%). La Tabla I muestra que la mayor dificultad se concentra en el pronóstico de la categoría normal. Por otro lado, el cluster 1 presenta, en líneas generales, mejores métricas que el cluster 2.

Asimismo, el HSS alcanza valores más altos (más favorables) para el pronóstico de la precipitación en el sector centro-oeste que en el sector este de la región de estudio. Finalmente, existe cierta superioridad en el desempeño de los pronósticos de OND respecto a los de EFM.

4) CONCLUSIONES

En este estudio se elaboraron modelos estadísticos para la predicción probabilística de la precipitación mensual en la CRB. Se exploraron diversos predictores de gran escala y escala regional, y finalmente se obtuvieron modelos de predicción basados en las técnicas de RLM, GAM, SVR, y ANN. La verificación mostró un mejor rendimiento determinístico para los modelos basados en ANN. Por otro lado, se hallaron métricas de verificación probabilística favorables particularmente para el sector identificado como cluster 1, y en los meses que componen a la primavera.

REFERENCIAS

- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., y Thépaut, J-N., 2019: ERA5 monthly averaged data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Acceso < 01-06-2021>), 10.24381/cds.f17050d7
- Mason, S.J. y Stephenson, D.B., 2008: How Do We Know Whether Seasonal Climate Forecasts are Any Good? En: Troccoli, A., Harrison, M., Anderson, D.L.T., y Mason, S. (eds.) Seasonal Climate: Forecasting and Managing Risk (pp. 205-234). Dordrecht: Springer Netherlands.
- Subsecretaría de Agricultura (Dirección Nacional de Agricultura), Estimaciones Agrícolas, disponible en <https://datosestimaciones.magyp.gob.ar/reportes.php?reporte=Estimaciones>. Consultado en noviembre de 2022.