

PREDICCIÓN DE LA PRECIPITACIÓN DE INVIERNO EN LA REGIÓN DEL COMAHUE USANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Laura D. Waisbein¹, Marcela H. González^{1,2,3}, Alfredo L. Rolla^{2,3,4}, Sabrina N. Ayala^{1,2,3}, Paula C. Oliveri¹, Eugenia M. Garbarini⁵, Maximiliano S. Vita Sánchez^{1,5}

Contacto: ldwaisbein@gmail.com

¹ Departamento de Ciencias de la Atmósfera y los Océanos (FCEyN, UBA).

² Centro de Investigaciones del Mar y la Atmósfera (CIMA) (CONICET – UBA).

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

⁴ Facultad de Ciencias Exactas y Naturales (UBA).

⁵ Servicio Meteorológico Nacional (SMN)

Palabras clave: pronóstico probabilístico, ensamble de modelos, aprendizaje automático.

1) INTRODUCCIÓN

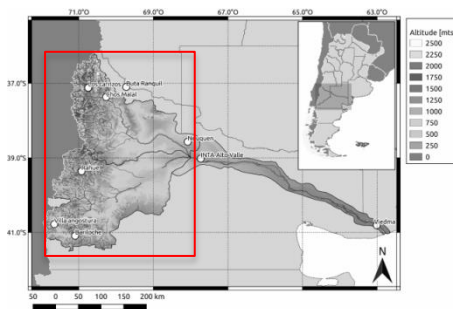


Figura 1: Región de estudio y estaciones utilizadas.

En la región del Comahue, en el noroeste de la Patagonia argentina, se encuentran las cuencas de los ríos Limay y Neuquén. Allí se localizan varias represas hidroeléctricas que aportan a la red eléctrica nacional. Las máximas precipitaciones tienen lugar durante el invierno y son las responsables principales del aumento del caudal. Es por ello que su pronóstico en escala estacional permite a los tomadores de decisiones estar preparados frente a situaciones extremas. Los pronósticos determinísticos en escala estacional no son precisos. Por tal motivo, se busca diseñar un pronóstico que utilice muchos modelos aplicando diferentes técnicas estadísticas para obtener un ensamble de modelos cuyo producto final sea un pronóstico probabilístico de precipitación. Este trabajo presenta un método de pronóstico estadístico para la precipitación de invierno (abril - septiembre) en las cuencas de los ríos Limay y Neuquén que incorpora técnicas de aprendizaje automático para sumarlas a otras más convencionales como la regresión lineal múltiple.

2) DATOS Y METODOLOGÍA

Para el trabajo se utilizaron datos mensuales de precipitación para el período 1981-2020 provenientes de diversas fuentes. A partir de esos datos se obtuvieron las series de precipitación acumulada trimestral para tres estaciones localizadas en la cuenca del Limay y cuatro en la cuenca del Neuquén (Figura 1), se calculó la precipitación media para cada subcuenca y con las series resultantes se entrenaron los modelos de pronóstico.

Para definir los predictores se utilizaron datos mensuales provenientes de reanálisis del NCEP de altura geopotencial (HGT) en 1000, 500 y 200 hPa, las componentes zonal y meridional del viento (u, v) en 850 hPa, el contenido de agua precipitable (tcw) y la temperatura de la superficie del mar (SST), considerando sólo aquellos predictores con significado físico. Luego se correlacionaron los predictores del mes previo al trimestre a pronosticar con la precipitación acumulada en el trimestre. Las áreas con correlación significativa se definieron como predictores. De todos ellos se eligió un conjunto de predictores independientes utilizando la metodología LASSO.

Las metodologías aplicadas (James et al., 2013) para generar los modelos incluyen regresión lineal múltiple, modelos aditivos generalizados, regresión de soporte vectorial y cuatro arquitecturas de redes neuronales.

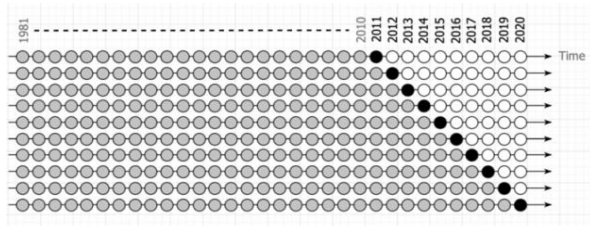


Figura 2: Esquema mostrando los períodos de entrenamiento y verificación.

El período de entrenamiento inicial fue 1981-2010 y la verificación fue realizada para el período 2011-2020, con la particularidad que para el pronóstico del año n los modelos fueron entrenados hasta el año $(n-1)$ (Figura 2). El ensamble se generó con los modelos que explicaban más del 50% de la varianza de la precipitación, es decir, con un coeficiente de determinación ajustado (R^2_{adj}) mayor

que 0,5:

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{(n - k - 1)} \quad (1)$$

donde n es el número de datos, k es el número de variables predictoras.

La evaluación del pronóstico probabilístico (Kumar et al., 2020) fue realizada mediante diagramas de confiabilidad y el cálculo del Brier score utilizando tres categorías: subnormal (BN), sobrenormal (AN) y normal (NOR) ya sea que el valor fuera inferior al primer tercil, superior al segundo tercil o comprendido entre el primer y segundo tercil, respectivamente. Están basados en el diagnóstico de pronósticos probabilísticos para un conjunto predefinido de eventos y así pueden ser construidos para cada categoría (BN, NOR y AN). El diagrama para las categorías individuales es útil para indicar si la calidad de los pronósticos depende de los resultados. Para cada probabilidad pronosticada, los diagramas de confiabilidad indican si el evento pronosticado ocurrió con la frecuencia pronosticada. El Brier score (BS) mide la precisión de las predicciones probabilísticas. Cuanto más bajo su valor para un conjunto de predicciones, mejor resultan las predicciones. Se define como el cuadrado de la mayor diferencia posible entre la probabilidad predicha y el resultado real:

$$BS = \frac{1}{n} \sum_{t=1}^N (f_t - o_t)^2 \quad (2)$$

donde f_t es la probabilidad pronosticada, o_t es el resultado real y n es el número de instancias de pronóstico.

3) RESULTADOS Y CONCLUSIONES

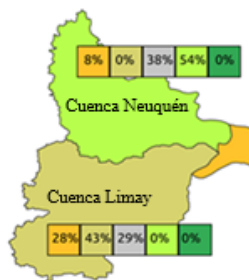


Figura 3: Pronóstico probabilístico ASO 2021

ASO 2021	LIMAY	NEUQUEN
Número de modelos	7	13
Pronosticado	43% de modelos con precipitación entre 230 y 254mm (Q2)	54% de modelos con precipitación entre 97 y 118mm (Q4)
Observado	Normal a inferior	Normal a superior

Tabla 1: Pronóstico de precipitación ASO 2021. El sombreado indica la categoría de precipitación: muy sub y subnormal (Q1, Q2) en naranja; normal (Q3) en gris; sobre y muy sobrenormal (Q4, Q5) en verde.

La Figura 3 muestra un ejemplo del pronóstico probabilístico de precipitación para el trimestre ASO 2021 para las cuencas de los ríos Limay y Neuquén y la Tabla I, el número de modelos del ensamble, el quintil (Q) pronosticado con mayor probabilidad de ocurrencia y la precipitación observada en el trimestre. El pronóstico para este trimestre demostró un buen desempeño en ambas subcuencas.

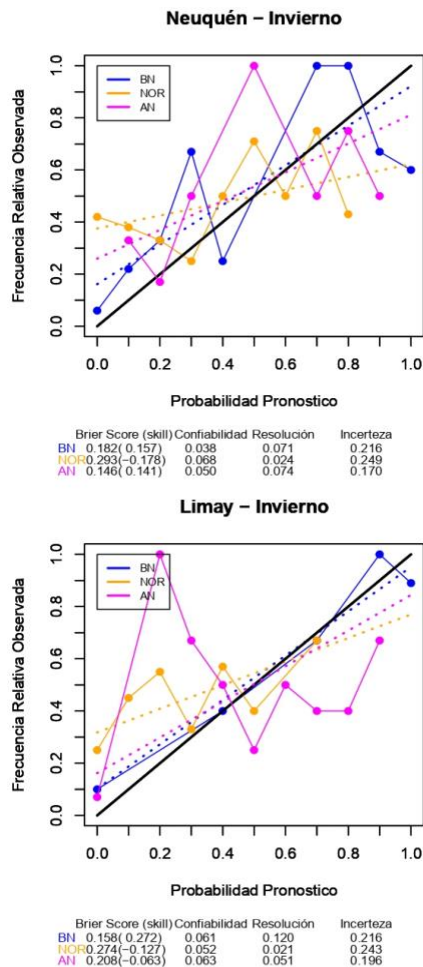


Figura 4: Diagramas de confiabilidad para Neuquén (panel superior) y Limay (panel inferior) para el invierno.

La Figura 4 muestra los diagramas de confiabilidad para el río Neuquén (panel superior) y para el río Limay (panel inferior) para el invierno. En cada gráfico se incluye la recta de regresión ponderada para cada categoría. El Brier Score se detalla debajo de cada gráfico. El Brier Skill Score se calcula usando pronósticos de referencia con probabilidades climatológicas (1/3). Los pronósticos tienden a sobreestimar la ocurrencia de valores reales cuando la probabilidad es alta y subestimarlos cuando la probabilidad es baja. El BS presenta mejores valores en las categorías BN y AN para ambas subcuencas mientras que el BSS para dichas categorías muestra un valor positivo indicando que el pronóstico probabilístico propuesto es mejor que el pronóstico de referencia (climatología) para la cuenca del Neuquén y para la categoría BN para la cuenca del Limay, en tanto que registra un valor levemente negativo para la categoría AN en esta cuenca donde se parecería más al pronóstico de referencia. Los diagramas de confiabilidad muestran que el pronóstico probabilístico es bastante eficiente en invierno, época en que se registran las mayores precipitaciones en las cuencas de los ríos Limay y Neuquén y que impactan en la generación de energía hidroeléctrica.

AGRADECIMIENTOS

Al SMN y la AIC por los datos. A los proyectos UBACYT 2020-2022 20620120100003ba y 2017-2019 20020160100009ba por la financiación.

REFERENCIAS

James, G., Witten, D., Hastie, T. y Tibshirani, R., 2013: An Introduction to Statistical Learning with Applications in R, Springer, NY, USA, DOI 10.1007/978-1-4614-7138-7.

Kumar, A., Ceron, J., Coelho, C., Ferranti, L., Graham, R., Jones, D., Merryfield, W., Muñoz, A., Pai, S., Rodriguez, E., 2020: Guidance on Operational Practices for Objective Seasonal Forecasting, World Meteorological Organization, WMO-No. 1246, 106 pp.