

Metodología de verificación del pronóstico estacional por consenso

Nota Técnica SMN 2021-110

María Mercedes Poggi¹, María de los Milagros Skansi¹, Ramón de Elía²

¹ Dirección Central de Monitoreo del Clima, Dirección Nacional de Ciencia e Innovación en Productos y Servicios

² Dirección Nacional de Ciencia e Innovación en Productos y Servicios.

Noviembre 2021

Información sobre Copyright

Este reporte ha sido producido por empleados del Servicio Meteorológico Nacional con el fin de documentar sus actividades de investigación y desarrollo. El presente trabajo ha tenido cierto nivel de revisión por otros miembros de la institución, pero ninguno de los resultados o juicios expresados aquí presuponen un aval implícito o explícito del Servicio Meteorológico Nacional.

La información aquí presentada puede ser reproducida a condición que la fuente sea adecuadamente citada.

Resumen

El proceso de verificación representa un componente esencial de la predicción climática, especialmente en aquellos pronósticos que son relevantes para diversos sectores. En el presente trabajo se presenta el procedimiento y las metodologías utilizadas para la verificación del pronóstico estacional por consenso elaborado por el Servicio Meteorológico Nacional de Argentina. La verificación también comprende la comparación entre el desempeño de dicho pronóstico con el de las previsiones de otros centros o modelos climáticos.

Abstract

The verification process represents an essential component of climate prediction, especially in those forecasts that are relevant to various sectors. In this work, the procedure and methodologies used to verify the seasonal consensus forecast of the National Meteorological Service of Argentina are presented. The verification also includes a comparison between the performance of this forecast with forecasts from other centers or climate models.

Palabras clave: Verificación, pronóstico estacional, métricas.

Citar como:

Poggi, M. M., Skansi M. M., de Elía R., 2021: Metodología de verificación del pronóstico estacional por consenso. Nota Técnica SMN 2021-110.

1. INTRODUCCIÓN

En los últimos años el Servicio Meteorológico Nacional (SMN) ha puesto una creciente atención en la necesidad de establecer verificaciones objetivas de todos los pronósticos que produce. Entre ellos, se encuentra el pronóstico estacional, el cual es utilizado por una gran variedad de usuarios y resulta de relevancia para aquellos sectores que necesitan tener algún tipo de predictibilidad para planificar sus acciones en la escala de mediano plazo, como es el caso de las actividades agrícolas, hídricas, el sector energético, entre otros. En esta Nota Técnica se presentan los esfuerzos de verificación de dicho pronóstico.

El pronóstico estacional del SMN consiste en un pronóstico probabilístico por consenso elaborado desde la Dirección Central del Monitoreo del Clima en conjunto con la Dirección Nacional de Servicios Sectoriales. También participan el Departamento de Ciencias de la Atmósfera y los Océanos de la Universidad de Buenos Aires, y el Centro de Investigaciones del Mar y la Atmósfera del Consejo Nacional de Investigaciones Científicas y Técnicas. Dicho pronóstico se realiza mensualmente para todo el país para las variables temperatura media y precipitación acumulada, y corresponde a un campo o mapa con un sistema de probabilidades de tres categorías (“superior a lo normal”, “normal” e “inferior a lo normal”). Tiene la particularidad de contener una componente subjetiva: su elaboración se basa en el análisis que realizan los expertos no sólo de las condiciones atmosféricas-oceánicas prevalentes, sino también de las salidas disponibles de diversos modelos climáticos. Si bien todos los expertos utilizan como guía o referencia la misma información, sus valoraciones son únicas y difieren entre sí producto de su experiencia y percepción.

El proceso de verificación tiene como principal objetivo la evaluación del pronóstico por consenso. A su vez, comprende la comparación entre el desempeño de dicho pronóstico con el de los pronósticos provenientes de otros centros, de modelos climáticos, o del propio SMN. El propósito de esto es determinar si la contribución subjetiva de experimentados pronosticadores podría mejorar las predicciones derivadas de algoritmos.

2. MODELOS PARA LA COMPARACIÓN

Las predicciones estacionales, ya sean globales o regionales, pueden ser obtenidas a partir de modelos dinámicos o empíricos (estadísticos), cada uno presentando diversas ventajas y desventajas a la hora de su implementación. También existen otros sistemas de pronósticos que se basan en la combinación de ambos; éstos son reconocidos como modelos dinámico-estadísticos.

Los pronósticos generados a partir del enfoque dinámico (o pronósticos dinámicos) utilizan la información de los modelos de circulación general (GCMs, por sus siglas en inglés), mientras que los pronósticos derivados del enfoque estadístico (o pronósticos estadísticos) se construyen mediante técnicas estadísticas que relacionan el clima estacional con ciertos predictores. Los pronósticos dinámico-estadísticos, por su parte, se basan en un post procesamiento estadístico de las salidas de los GCMs.

Para el análisis comparativo se consideraron pronósticos estacionales generados a partir de todos los enfoques. En relación a los enfoques dinámico y dinámico-estadístico, los pronósticos corresponden al del Ensamble Multi-Modelo Norteamericano (NMME, por sus siglas en inglés) y a diversas calibraciones estadísticas del mismo realizadas por otros centros o institutos.

Los modelos utilizados son los siguientes:

- **NMME (North American Multi-model Ensemble):** El NMME (Kirtman y otros, 2014) es un sistema de predicción estacional multimodelo que combina los pronósticos de los principales GCMs de Estados Unidos (incluidos NOAA, NCAR y NASA), y del Environment and Climate Change Canada. Generalmente, dicho sistema presenta una mayor habilidad en comparación con los modelos individuales que lo componen (Palmer y otros, 2004; Jin y otros, 2008; entre otros). Los modelos activos del NMME (Fase II, versión 1) se listan en la Tabla 1.

Tabla 1: Modelos que actualmente componen la Fase II del NMME, con los organismos que lo desarrollan.

Modelos	Organismo
NCEP CFSv2	NOAA National Centers for Environmental Prediction
GFDL SPEAR	NOAA Geophysical Fluid Dynamics Laboratory
NASA GEOS2S	NASA Global Modeling and Assimilation Office
ECCC CanCM4i	Environment and Climate Change Canada
ECCC GEM-NEMO	
NCAR/RSMAS CCSM4	National Center for Atmospheric Research

Los datos y figuras de los pronósticos del NMME se obtienen, respectivamente, a partir de los siguientes enlaces: <ftp://ftp.cpc.ncep.noaa.gov/NMME/prob/netcdf/>, <https://www.cpc.ncep.noaa.gov/products/NMME/prob/pac/>.

Mayor información y descripción del NMME, tanto de la Fase I como de la Fase II, se puede encontrar en: https://www.cpc.ncep.noaa.gov/products/NMME/about_test.html.

- **CLIMAX:** Los modelos del proyecto CLIMAX (Osman y otros, 2021) se basan en la calibración de las salidas de cada uno de los modelos que integran el NMME, a excepción del GFDL SPEAR (operativo desde febrero de 2021). Se utiliza la técnica de regresión de ensamble (Unger y otros, 2009) para calibrar cada modelo individual. Luego, estos modelos calibrados se combinan de diversas maneras para crear seis pronósticos probabilísticos multi-modelo distintos, en función del peso atribuido a cada modelo individual. En particular, el pronóstico operativo corresponde a la variante “wsereg-mean_cor”, que considera una ponderación proporcional a la correlación media entre las predicciones y las observaciones.
- **IRI (International Research Institute for Climate and Society):** Desde abril de 2017, el pronóstico probabilístico del IRI se basa en la calibración de las salidas de cada uno de los modelos que integran el NMME, a excepción del GFDL SPEAR. Se emplea la regresión logística extendida (Wilks, 2009) para calibrar y generar los pronósticos de probabilidad de cada modelo. Luego, estas probabilidades se promedian con igual peso para crear el pronóstico probabilístico multi-modelo (Vigaud y otros, 2017).

Los datos y figuras de los pronósticos del IRI se obtienen, respectivamente, a partir de los siguientes enlaces: https://iridl.ideo.columbia.edu/SOURCES/IRI/FD/NMME_Seasonal_Forecast/, <https://iri.columbia.edu/our-expertise/climate/forecasts/seasonal-climate-forecasts/>.

Mayor información y descripción de la metodología utilizada por el IRI se puede encontrar en: <https://iri.columbia.edu/our-expertise/climate/forecasts/seasonal-climate-forecasts/methodology/>. La metodología empleada previo a abril de 2017 se detalla en Barnston y otros (2010), y en el enlace: <https://iri.columbia.edu/our-expertise/climate/forecasts/iri-real-time-seasonal-climate-forecasts-and-models/>.

- **SMN-CPT (Climate Predictability Tool):** El SMN-CPT corresponde a un modelo estadístico, dado que no aplica información proveniente de ningún GCMs. Emplea la Herramienta de Predicción Climática (CPT, por sus siglas en inglés) (Mason y otros, 2020) desarrollada por el IRI y la técnica de análisis de

correlación canónica (Barnston, 1994) para realizar predicciones a nivel país y por regiones. El modelo se entrena en el período 1961-2018 y utiliza datos de reanálisis de diversas variables atmosféricas en altura (altura geopotencial en 500 hPa, viento zonal en 250 hPa, viento meridional en 850 hPa y velocidad vertical en 700 hPa) y la temperatura superficial del mar como predictores. Mayor información del CPT puede encontrarse en el siguiente enlace: <https://iri.columbia.edu/our-expertise/climate/tools/cpt/>.

- **SMN-CPT.NMME:** Constituye una variación del modelo SMN-CPT, combinando el modelo estadístico con el resultado de la calibración de la media del ensamble del NMME. En este caso, el período de entrenamiento se limita a los años del período retrospectivo del NMME (1982-2010).

Una mayor explicación del SMN-CPT y SMN-CPT.NMME se encontrará en la nota técnica de Herrera y otros (en proceso).

3. DATOS

Los mapas de pronósticos por consenso se realizan por áreas o zonas climáticas. A partir de ellos, se obtienen las predicciones en puntos de estación. Se consideran 86 estaciones individuales no distribuidas de manera uniforme en el territorio argentino, que se presentan en la Figura 1.

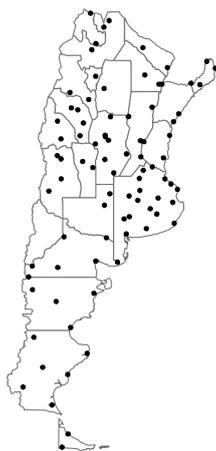


Fig. 1: Distribución de las 86 estaciones meteorológicas del SMN consideradas por el pronóstico por consenso.

Para la verificación se utilizan las observaciones en las mismas estaciones, de las que además se cuenta con información de los terciles inferior y superior de las variables temperatura media y precipitación acumulada. La información de estos umbrales es necesaria para establecer si una observación pertenece a la categoría “normal”, “superior a lo normal” o “inferior a lo normal”, y luego poder compararla con lo pronosticado.

3.1 Datos de los pronósticos NMME, CLIMAX e IRI

Los modelos NMME, CLIMAX e IRI no presentan datos en estaciones individuales, sino que proveen información en todos los puntos de retícula de 1° de latitud por 1° de longitud comprendidos en Argentina. Más aún, en las figuras que ofrecen a los usuarios, las mismas que analizan los expertos al momento de realizar el pronóstico por consenso, sólo se observa la categoría más probable (con una escala de probabilidades cada 5% que inicia a partir del 40%) en donde el océano está enmascarado. Es por ello que, previo a realizar cualquier análisis o evaluación, se les aplican dos máscaras a los datos crudos de dichos

modelos: la primera corresponde al océano y la segunda, a aquellos puntos de retícula con probabilidades dominantes (redondeadas al múltiplo de 5% más cercano) inferiores al 40%. El objetivo de esto último es retener los pronósticos informativos o no-climatológicos, es decir, sólo preservar aquella información que difiere sustancialmente de la predicción climatológica. La predicción climatológica es aquella que asigna igual probabilidad de ocurrencia a cada uno de los terciles. El empleo de máscaras y la manipulación de los datos crudos permiten reproducir las imágenes brindadas a los usuarios (Figura 2), así como también generar una escala de color unificada para todas las figuras de una misma variable.

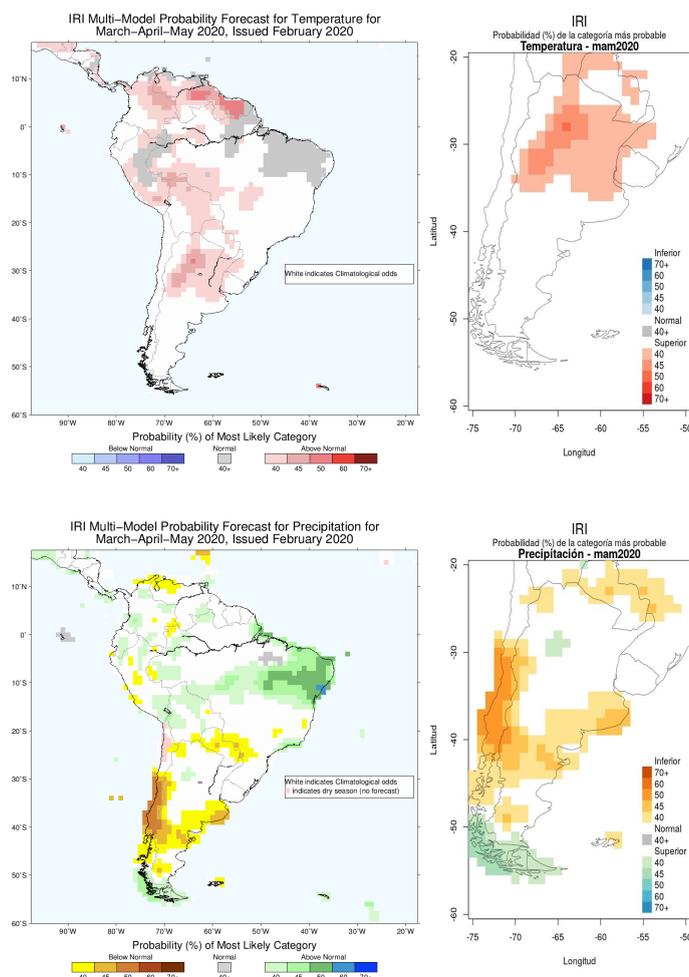


Fig. 2: Campos del pronóstico probabilístico estacional del IRI para la temperatura (arriba) y precipitación (abajo) para el trimestre marzo-abril-mayo de 2020, y reproducciones de los mismos aplicando diversas máscaras a los datos.

La máscara aplicada al océano pertenece al modelo NMME y puede descargarse desde el siguiente enlace: <https://iridl.ldeo.columbia.edu/SOURCES/Models/NMME/LSMASK/>. Su utilización es particularmente importante en el caso de los pronósticos de temperatura, dado que éstos sólo son válidos sobre el continente.

Una vez realizado esto, se interpola la información para obtener predicciones en cada punto de estación. Se considera la interpolación por el vecino más próximo. Cabe destacar que aplicar máscaras ocasiona que la cantidad de puntos de retícula y, por lo tanto, de estaciones con pronóstico informativo sea considerablemente menor (dado que muchas áreas resultan enmascaradas).

3.2 Datos de los modelos SMN-CPT y SMN-CPT.NMME

Los modelos SMN-CPT y SMN-CPT.NMME sí pronostican en puntos de estación pero lo hacen en una menor cantidad de estaciones que el pronóstico por consenso. Consideran 79 estaciones para la lluvia (Figura 3 izq.) y 67 para la temperatura (Figura 3 centro der.). Las regiones o áreas climáticas que proponen también son distintas. Para la precipitación, plantean 6 regiones (Figura 3 centro izq.) mientras que para la temperatura, 12 (Figura 3 der.). Para estos modelos no fue necesario realizar una interpolación.

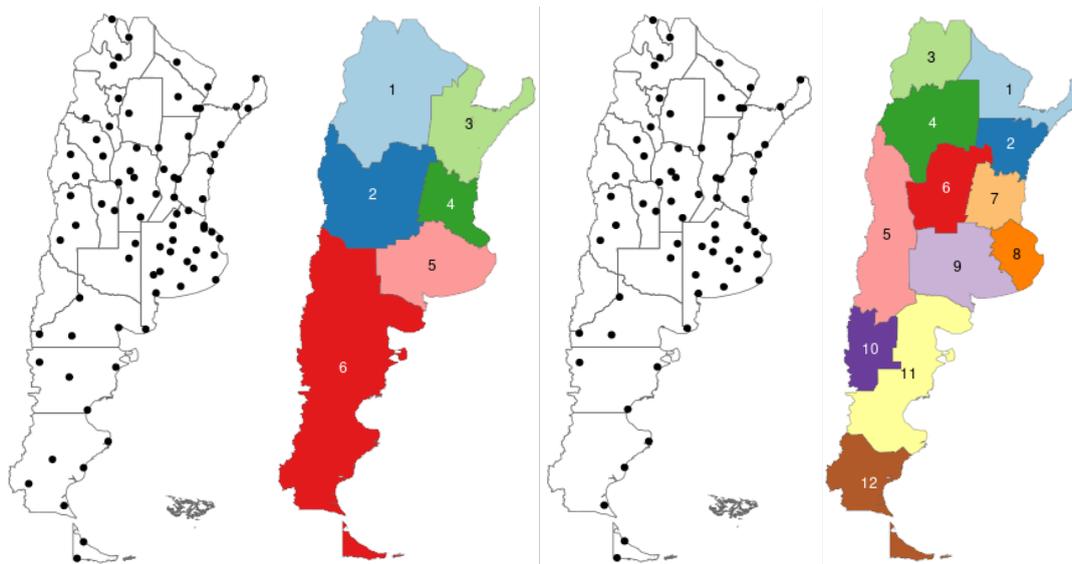


Fig. 3: Campos de distribución de las estaciones meteorológicas del SMN y de las regiones consideradas por los modelos SMN-CPT y SMN-CPT.NMME para el pronóstico de la precipitación (izq. y centro izq.) y temperatura (der. y centro der.).

4. PERÍODO DE ANÁLISIS

El período de estudio abarca desde el trimestre enero-febrero-marzo de 2018 hasta la fecha. Dicho período cuenta, en forma ininterrumpida, con datos asociados al pronóstico por consenso. En cuanto a los otros modelos, no todos presentan datos en todos los trimestres. Los pronósticos evaluados del SMN-CPT y del SMN-CPT.NMME, por ejemplo, comienzan en el 2019 porque a partir de ese año se consideró una nueva regionalización y se fijaron los predictores para cada región/trimestre. El CLIMAX, por su parte, tiene una discontinuidad a principios del 2021 debido a un cambio de dos de los modelos que integran el NMME. Del sistema multimodelo no se encontraron los datos de libre acceso del 2018. La Figura 4 resume la disponibilidad de datos de precipitación de todos los modelos considerados.

PRECIPITACIÓN

	AÑO/TRIMESTRE	EFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDE	DEF
CONSENSO	2018												
	2019												
	2020												
	2021												
NMME	2018												
	2019												
	2020												
	2021												
CLIMAX	2018												
	2019												
	2020												
	2021												
IRI	2018												
	2019												
	2020												
	2021												
SMN-CPT	2018												
	2019												
	2020												
	2021												
SMN-CPT. NMME	2018												
	2019												
	2020												
	2021												

Fig. 4: Disponibilidad de datos de pronóstico de precipitación de todos los modelos considerados, desde el trimestre enero-febrero-marzo de 2018 hasta abril-mayo-junio de 2021. Los colores verdes corresponden a trimestres con datos mientras que los colores rojos, a trimestres sin información.

En los modelos SMN-CPT y SMN-CPT.NMME, los pronósticos de precipitación de los trimestres junio-julio-agosto no cuentan con información en el norte y noroeste argentino porque incluyen una máscara para la estación seca.

La Figura 5 resume la disponibilidad de datos de temperatura de todos los pronósticos considerados.

TEMPERATURA

	AÑO/TRIMESTRE	EFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDE	DEF
CONSENSO	2018												
	2019												
	2020												
	2021												
NMME	2018												
	2019												
	2020												
	2021												
CLIMAX	2018												
	2019												
	2020												
	2021												
IRI	2018												
	2019												
	2020												
	2021												
SMN-CPT	2018												
	2019												
	2020												
	2021												
SMN-CPT. NMME	2018												
	2019												
	2020												
	2021												

Fig. 5: Ídem Figura 4 pero para la variable temperatura.

Cabe destacar que, por trimestre, sólo se consideran para el análisis aquellos modelos que presentan pronósticos no-climatológicos en alguna región de Argentina. Por ejemplo, para el trimestre marzo-abril-mayo de 2019, el modelo IRI, si bien cuenta con datos, no exhibe pronóstico informativo (Figura 6). Esto se debió a que las probabilidades de las categorías dominantes fueron inferiores al 40% y, por lo tanto, resultaron enmascaradas.

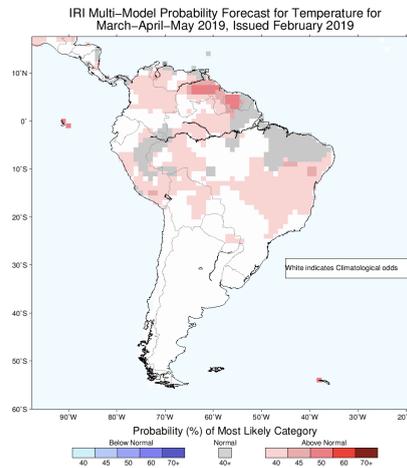


Fig. 6: Mapa del pronóstico probabilístico estacional del IRI para la temperatura y el trimestre marzo-abril-mayo de 2019. Notar la ausencia de pronóstico informativo.

5. REGIONES DE ESTUDIO

El estudio de verificación fue realizado en toda Argentina y por áreas, para también evaluar el desempeño del pronóstico en las distintas zonas climáticas. Se consideraron once regiones determinadas por el pronóstico por consenso, presentadas en la Figura 7.



Fig. 7: Delimitación de las once regiones utilizadas para la verificación por zonas climáticas.

6. MÉTRICAS PARA LA VERIFICACIÓN

Para la evaluación del desempeño del pronóstico por consenso se utilizaron diversas métricas sugeridas por la “Guía de Verificación de los Pronósticos Climáticos Estacionales Operativos” (Mason, 2018), de la Organización Meteorológica Mundial (OMM). Los procedimientos que se llevaron a cabo se enmarcan dentro del “Plan transversal de verificación de pronósticos del SMN” (de Elia, 2021), e intentan ser coherentes con los adoptados en las verificaciones de los otros pronósticos de la institución.

En su texto, Mason (2018) describe y recomienda los pasos a seguir para la verificación de los pronósticos probabilísticos estacionales, que pueden ser resultados de modelos, la evaluación subjetiva de expertos o una combinación de ambos. Allí diferencia los procedimientos a seguir según si se tiene una serie de pronósticos o un mapa de pronóstico individual para un trimestre o temporada específica.

Las métricas sugeridas a emplear cuando se dispone de un solo campo de pronóstico de un trimestre en particular son las siguientes:

- *Hit (Heidke) score (para las categorías con mayor, segunda mayor y menor probabilidad)*
- *Ignorance score*
- *Average interest rate*

También se aconseja incluir el campo observado de precipitación o temperatura, pero utilizando el sistema de categorías. No se deben usar mapas de anomalías o de desvíos porcentuales del promedio, porque en ninguno de estos casos queda claro en qué categoría se encuentra la observación. Se deben construir mapas en donde se sombreen las áreas con mediciones dentro el rango normal, o por encima o por debajo de éste.

Algunas de las métricas de verificación recomendadas a utilizar cuando se cuenta con una serie de pronósticos son las siguientes:

- *Diagramas y área ROC (Relative operating characteristics) (para las categorías con mayor, segunda mayor y menor probabilidad)*
- *Diagrama de confiabilidad (para las categorías con mayor, segunda mayor y menor probabilidad)*

Según Mason (2018), estas métricas, con excepción del *Average interest rate*, constituyen el conjunto mínimo de procedimientos que todos los centros operativos de pronóstico deben esforzarse por calcular.

Se puede observar que la verificación de un pronóstico para un trimestre individual en varios puntos del país requiere de procedimientos distintos de aquellos a considerar cuando se tiene una serie de predicciones para un mismo punto (dado que el análisis en el espacio y el tiempo son esencialmente diferentes, en particular por la alta correlación espacial). Si bien un sistema de pronóstico sólo puede ser evaluado adecuadamente sobre la base de una larga serie, analizar el desempeño por trimestre también resulta útil y valioso, en particular para ser conscientes del valor de la información brindada al público en un determinado período.

Cabe destacar que el conjunto de métricas para la verificación fue aplicado no sólo al consenso, sino también a los pronósticos de todos los modelos considerados, tanto para evaluar sus desempeños como para utilizar sus resultados a modo comparativo.

6.1 Referencias de métricas

Las métricas calculadas para la verificación de los mapas de pronóstico probabilístico para un trimestre en particular son:

- **Hit scores (HS):** Los *HS* establecen con qué frecuencia (en porcentaje) se observan las categorías pronosticadas de mayor, menor y segunda mayor probabilidad. Estas métricas son las más populares y pueden ser utilizadas como punto de partida para la verificación por su fácil interpretación y comunicación. Sin embargo, para su cálculo, no ponderan el valor de la probabilidad, sólo contabilizan si se acierta o no en la categoría predicha. Una mayor referencia de estas métricas puede encontrarse en Mason (2012).

En ocasiones, puede ocurrir que dos categorías presenten igual y máxima probabilidad en el pronóstico. Por ejemplo, en un trimestre y ubicación dada, puede ocurrir que el pronóstico atribuya un 40% de probabilidad a las categorías “superior a lo normal” y “normal”. Si efectivamente se observa una de ellas, Mason (2018) recomienda no contabilizar un acierto completo, sino medio. Esta recomendación no se tuvo en cuenta a la hora de calcular los *HS* porque puede generar confusión a la hora de presentar y comunicar los resultados.

- **Ignorance score (IS):** El *IS* (Roulston and Smith, 2002) mide, dado un determinado pronóstico, cuánta información adicional se necesita para lograr la categoría observada. A diferencia de los *HS*, sí considera el valor de la probabilidad de la categoría de acierto, ponderando las mayores probabilidades, y se calcula a partir de su logaritmo en base 2. El índice varía de 0 para un pronóstico perfecto hasta infinito para un pronóstico completamente malo. Si bien es la métrica más apropiada, su interpretación y comunicación no es intuitiva, y es poco utilizada en las verificaciones de otros tipos de pronósticos.
- **Average interest rate (AIR):** El *AIR* mide cuál es la tasa de ganancia cuando se invierte en un pronóstico específico, en comparación con invertir en la climatología. Al igual que el *IS*, considera para su cálculo el valor de probabilidad de la categoría de acierto, dividiéndola por la probabilidad climatológica. Dicha métrica es positiva para pronósticos “buenos” (mejores que la climatología). Para pronósticos perfectos en un sistema de tres categorías, su máximo valor es de 200%, mientras que para pronósticos totalmente malos, su mínimo es de -100%. Una mayor explicación del *AIR* se presenta en Hagedorn y Smith (2008). Esta métrica tampoco es de fácil interpretación y comunicación, y es poco utilizada en las verificaciones de otros tipos de pronósticos.

Algunas de las métricas consideradas para la verificación de una serie de pronósticos probabilísticos son:

- **Curva ROC:** La *curva ROC* determina, gráficamente, la habilidad del pronóstico para discriminar entre eventos y no eventos. Para ello, se grafica la tasa de aciertos o *HR* (por sus siglas en inglés) de una predicción versus la tasa de falsas alarmas o *POFD* (por sus siglas en inglés). Estos valores se obtienen a partir de una tabla de contingencia construida para cada evento o categoría y dado un umbral de probabilidad. Mayor información sobre la *curva ROC* puede encontrarse en Harvey y otros (1992).
- **Área ROC:** El *área ROC* mide la habilidad del pronóstico en discriminar un evento de un no evento. Es precisamente el área bajo la *curva ROC*. Un *área ROC* superior a 0.5 refleja una discriminación positiva, siendo 1 su puntaje máximo. Un mayor detalle de esta métrica se presenta en Hogan y Mason (2012).
- **Diagrama de confiabilidad:** El *diagrama de confiabilidad* (Hartmann y otros, 2002) permite determinar la confiabilidad y la resolución de las predicciones. Establece con qué frecuencia ocurre una probabilidad de pronóstico. Para ello, se grafica la frecuencia observada versus la probabilidad del pronóstico para un determinado evento. Puede ser construido para cada categoría por separado o para todas juntas.

Cabe destacar que para el cálculo de todas las métricas y los intervalos de confianza (posteriormente presentados) sólo se consideran las estaciones meteorológicas que poseen conjuntamente pronóstico y

observación. Pueden existir estaciones sin pronóstico debido a: empleo de máscaras (en el caso de los modelos NMME, CLIMAX e IRI); trimestre con estación seca para la variable precipitación; o pronóstico de climatología tanto para la temperatura como para la precipitación. También puede ocurrir la ausencia de observaciones para la posterior verificación. Frente a esto, para una correcta interpretación de las verificaciones a la hora de la comparación, junto con los resultados de las métricas, se añade la información del número de predicciones intervinientes en cada caso.

Un ejemplo clarificador de estas consideraciones se puede encontrar en el siguiente enlace: https://www.cpc.ncep.noaa.gov/products/predictions/90day/skill_exp.html. Aquí se presentan algunos resultados de la verificaciones del pronóstico estacional de precipitación y temperatura elaborado por Centro de Predicción Climática (CPC) de la NOAA, para los inviernos boreales del período 1995-1996 hasta 2014-2015 y los Estados Unidos. La métrica de verificación utilizada se calcula en aquellas áreas que no poseen pronóstico de climatología. En las figuras de los resultados se incluye, además, un "porcentaje de cobertura" referido al porcentaje del país con pronóstico no-climatológico.

En el Apéndice A se presentan casos de estudio para ilustrar los pasos a seguir para calcular las métricas recomendadas y construir los gráficos. También se incluyen ecuaciones y descripciones simples.

6.2 Intervalos de confianza

Dado que se cuenta con una muestra pequeña de pronósticos estacionales, los resultados de las métricas acarrearán incertidumbres, siendo estimadores no del todo precisos. Una forma de subsanar estas incertidumbres es calculando los intervalos de confianza mediante la técnica de remuestreo o bootstrap (Efron, 1979) con reposición.

Los intervalos de confianza consisten en un rango de valores, definidos por sus límites superior e inferior, entre los cuales, con una probabilidad definida por el nivel de confianza, se encuentra el valor real de un parámetro poblacional a estimar; en este caso, el parámetro está representado por el verdadero valor de una dada métrica de verificación. Existen diferentes métodos para el cálculo de los intervalos de confianza mediante la técnica bootstrap. Aquí se consideró el método de los percentiles y un nivel de confianza del 90%. Si bien existe otro método que genera resultados más precisos y corrige además los sesgos, su complejidad matemática impide su fácil entendimiento, transmisibilidad e implementación operativa, lo que lo hace menos atractivo para los propósitos pretendidos. Esto está en línea con el "Plan transversal de verificación de pronósticos del SMN" (de Elia, 2021), en donde se establece que, aún si las métricas y los índices son para consumo interno de la institución, es importante adoptar criterios simples. Al abordar un programa de verificación, es preferible utilizar medidas sencillas y fáciles de entender a implementar metodologías complejas (WMO, 2000).

En el Apéndice B se incorpora una mayor información de los intervalos de confianza y del método bootstrap.

Para la realización de todos los cálculos se utilizó como herramienta el R (R Core Team, 2021), un lenguaje de programación con enfoque al análisis estadístico. En particular, para el cómputo del ROC score y su curva asociada se empleó el paquete "verification" (NCAR, 2015).

7. RESULTADOS

En esta sección se presentan los resultados de la verificación de las series de pronósticos para todos los modelos y desde el trimestre enero-febrero-marzo de 2018 hasta abril-mayo-junio de 2021. Cabe destacar que del NMME, SMN-CPT y SMN-CPT.NMME sólo se tiene información desde el año 2019.

7.1 Precipitación

En la Figura 8 se muestra, para cada modelo verificado, la serie temporal de *HS* o de porcentaje de aciertos en las predicciones más probables, mientras que en la Figura 9, la serie del número de pronósticos informativos o no-climatológicos disponibles, en porcentaje.

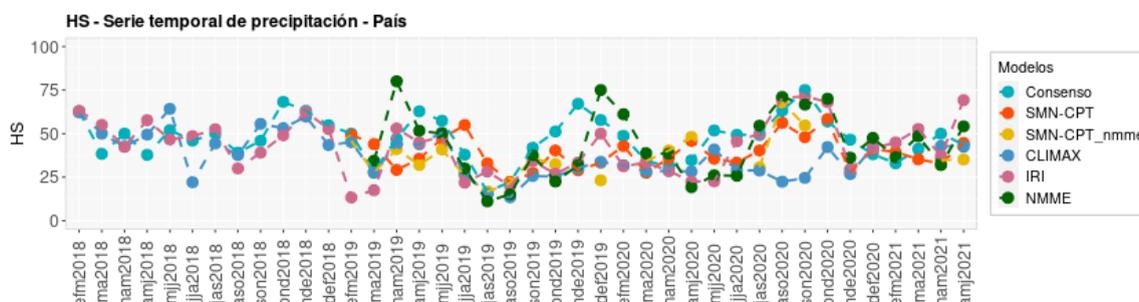


Fig. 8: Serie temporal de *HS* o de porcentaje de aciertos en las predicciones más probables para la precipitación y cada uno de los modelos considerados. El cálculo es a nivel país y abarca desde el trimestre enero-febrero-marzo de 2018 hasta abril-mayo-junio de 2021.

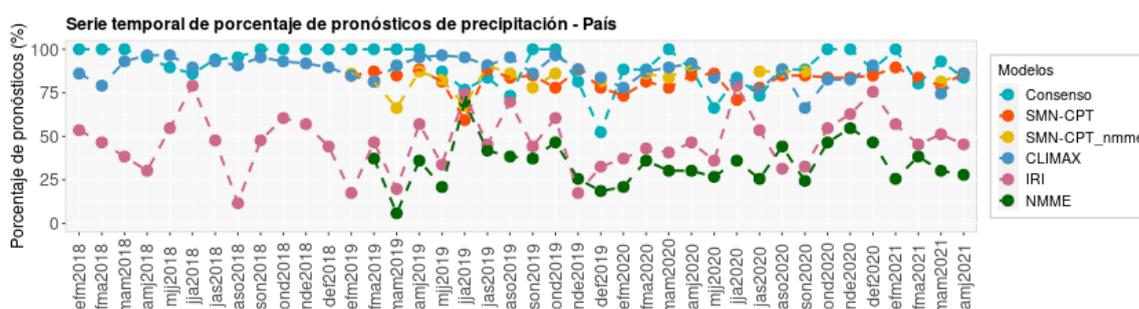


Fig. 9: Serie temporal de porcentaje de pronósticos informativos o no-climatológicos para la precipitación y cada uno de los modelos considerados. El cálculo es a nivel país y abarca desde el trimestre enero-febrero-marzo de 2018 hasta abril-mayo-junio de 2021.

A priori, se observa que ningún pronóstico se destaca por sobre otro según el *HS* en la categoría más probable. El IRI y el NMME sí resaltan por la poca cantidad de pronósticos que presentan en todos los trimestres considerados.

Tabla 2: Frecuencia relativa de aciertos en los pronósticos con mayor, segunda mayor y menor probabilidad para la precipitación, cada uno de los modelos considerados y el período de estudio.

<i>HS</i>	Modelos	Frecuencia relativa de aciertos en los pronósticos		
		Más probables	Segundo más probables	Menos probables
Precipitación	Consenso	0.47	0.27	0.26
	SMN-CPT	0.39	0.39	0.22
	SMN-CPT.NMME	0.36	0.40	0.23
	CLIMAX	0.38	0.36	0.26
	IRI	0.43	0.51	0.07
	NMME	0.41	0.45	0.14

En la Tabla 2 se resume la información proporcionada por la Figura 8. En ella se precisa, para cada modelo y la totalidad del período, la frecuencia relativa de veces que la precipitación observada perteneció a la categoría pronosticada con la primera, segunda y tercera mayor probabilidad.

A partir de la Tabla 2 se puede determinar que el consenso “gana” en la frecuencia con que se observan las categorías pronosticadas con mayor probabilidad. En los demás modelos, a excepción del CLIMAX, hay una máxima frecuencia de ocurrencias de los pronósticos con segunda mayor probabilidad. En CLIMAX, la cantidad promedio de aciertos en las predicciones con mayor y segunda mayor probabilidad es similar. También se destaca el IRI con una muy baja frecuencia de observaciones en los pronósticos menos probables.

La Tabla 3 resume los resultados promedios de las métricas *IS* y *AIR*, en función de los modelos y el total del período de estudio.

Tabla 3: Resultados de las métricas *IS* y *AIR* para la precipitación, cada uno de los modelos considerados y el período de estudio.

AIR/IS	Modelos	IS	AIR
Precipitación	Consenso	1.59	0.03
	SMN-CPT	Inf	0.04
	SMN-CPT.NMME	Inf	0.01
	CLIMAX	Inf	0.03
	IRI	1.56	0.04
	NMME	1.57	0.03

El *IS* determina cuán ignorante es el pronóstico en relación con la climatología. A partir de la Tabla 3 se ven resultados similares para el consenso, el NMME y el IRI. Para SMN-CPT, SMN-CPT.NMME y CLIMAX, el *IS* es infinito, el peor valor posible para esta métrica. Esto ocurre porque el *IS* es muy sensible a los malos pronósticos, es decir a aquellos que asignan una muy baja probabilidad a una categoría que finalmente sucede. Una vez que la métrica alcanza el infinito, éste no puede disminuirse, aún si el mal pronóstico es único y el resto de las veces, perfecto. En este caso, los valores de infinito se deben a determinados pronósticos en los trimestres: julio-agosto-septiembre de 2019 y diciembre-enero-febrero de 2020 para el SMN-CPT; agosto-septiembre-octubre de 2019 para el SMN-CPT.NMME; y mayo-junio-julio de 2019, junio-julio-agosto de 2019 y agosto-septiembre-octubre de 2019 para el CLIMAX.

Tabla 4: Áreas ROC para las categorías “superior a lo normal”, “normal” e “inferior a lo normal” para la precipitación, cada uno de los modelos considerados y el período de estudio.

Área ROC	Modelos	Categorías		
		Superior a lo normal	Normal	Inferior a lo normal
Precipitación	Consenso	0.60	0.50	0.56
	SMN-CPT	0.56	0.58	0.56
	SMN-CPT.NMME	0.50	0.59	0.50
	CLIMAX	0.60	0.52	0.56
	IRI	0.60	0.49	0.57
	NMME	0.54	0.53	0.51

Para el AIR, se observan valores similares para todos los modelos. Estos resultados determinan que, si se invierte en un determinado modelo en lugar de la climatología, se obtendrían ganancias del 3% para el consenso, el NMME y el CLIMAX; del 4% para el SMN-CPT y el IRI; y sólo del 1% para el SMN-CPT.NMME.

Los resultados de las áreas ROC se muestran en la Tabla 4. Comparando las áreas ROC de los modelos para las distintas categorías, se observa que el IRI y, en segundo lugar, el consenso y el CLIMAX presentan las mejores habilidades para distinguir las lluvias estacionales superiores e inferiores a lo normal. Los dos primeros modelos, sin embargo, no logran discriminar los eventos de precipitación en la categoría “normal”; el SMN-CPT y el SMN-CPT.NMME son los que exhiben la mejor performance en este caso.

7.2 Temperatura

Análogo a la Figuras 8 y 9, en la Figuras 10 y 11 se muestran las series temporales de HS en las predicciones más probables y de porcentaje de pronósticos informativos o no-climatológicos disponibles, respectivamente, pero para la variable temperatura.

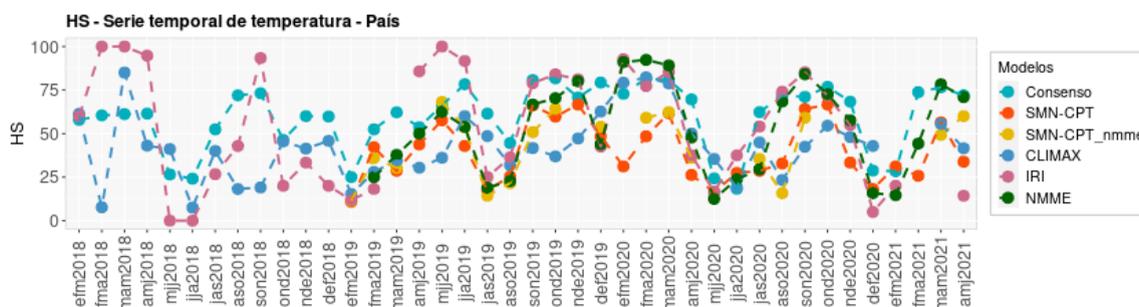


Fig. 10: Ídem Figura 8 pero para la temperatura.

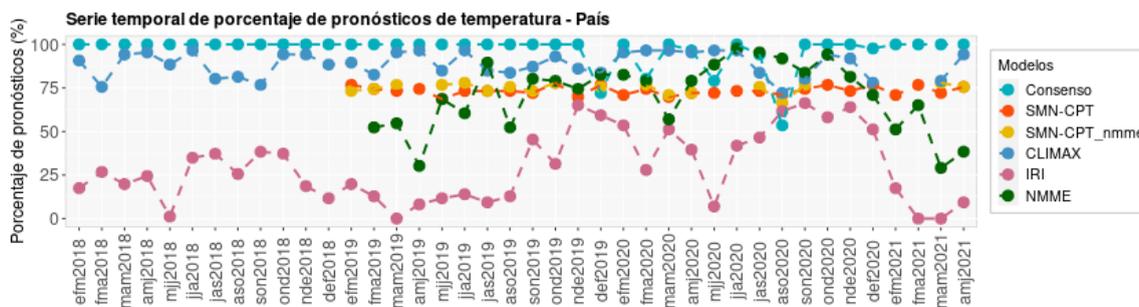


Fig. 11: Ídem Figura 9 pero para la temperatura.

Se destaca el IRI por la mínima cantidad de pronósticos que presenta en todos los trimestres considerados. Esto debe ser tenido en cuenta a la hora de formular conclusiones. Por ejemplo, en marzo-abril-mayo de 2018 el IRI exhibió casi un 100% de aciertos en las predicciones más probables. Sin embargo, sólo pronosticó en menos del 25% de las estaciones.

En la Tabla 5 se presenta, para cada modelo y la totalidad del período considerado, la frecuencia relativa de veces que la temperatura observada correspondió a la categoría pronosticada con la primera, segunda y tercera mayor probabilidad.

Tabla 5: Ídem Tabla 2 pero para la temperatura.

HS	Modelos	Frecuencia relativa de aciertos en los pronósticos		
		Más probables	Segundo más probables	Menos probables
Temperatura	Consenso	0.60	0.25	0.16
	SMN-CPT	0.40	0.38	0.23
	SMN-CPT.NMME	0.44	0.38	0.18
	CLIMAX	0.43	0.36	0.21
	IRI	0.59	0.37	0.04
	NMME	0.52	0.36	0.11

Se distinguen resultados más alentadores que en el caso de la precipitación. Nuevamente, el consenso presenta la mayor frecuencia relativa de aciertos en los pronósticos más probables, mientras que el IRI, la menor frecuencia relativa de aciertos en las predicciones con menor probabilidad. Para todos los modelos, los pronósticos con más alta probabilidad ocurren la mayoría de las veces; esto no se observaba en el caso de la lluvia.

La Tabla 6 resume los resultados promedios de las métricas *IS* y *AIR*, en función de los modelos y el total del período de estudio.

Tabla 6: Ídem Tabla 3 pero para la temperatura.

AIR/IS	Modelos	IS	AIR
Temperatura	Consenso	1.49	0.10
	SMN-CPT	Inf	0.10
	SMN-CPT.NMME	1.59	0.15
	CLIMAX	1.59	0.08
	IRI	1.47	0.10
	NMME	1.46	0.12

Según el *IS*, el NMME muestra los mejores resultados. Un *IS* de infinito se obtiene para el SMN-CPT debido a que, en los trimestres octubre-noviembre-diciembre de 2019, junio-julio-agosto de 2020 y agosto-septiembre-octubre de 2020, se pronosticó con una mínima probabilidad una categoría finalmente observada.

En función del *AIR*, el mejor modelo resulta ser el SMN-CPT.NMME. Si se apuesta por el SMN-CPT.NMME en lugar de la climatología, se obtendrían ganancias del 15% de la inversión original. Por el contrario, si se apuesta por el consenso, las ganancias serían del 10%. Nuevamente, los resultados son más alentadores que en el caso de la precipitación.

Los resultados de las *áreas ROC* se presentan en la Tabla 7. Según el *área ROC*, el consenso muestra los mejores resultados para las categorías “superior a lo normal” e “inferior a lo normal”, pero no así para el rango “normal”, si bien presenta alguna habilidad, contrario a lo sucedido con la precipitación. El SMN-CPT.NMME es el que mejor discrimina las condiciones normales de las no normales, seguido por el SMN-CPT. Esto ya se había observado para el caso de la lluvia. Cabe destacar que todos los modelos tienen una menor performance en distinguir los eventos fríos, en comparación con los normales y los cálidos. Esto podría deberse a que, en un contexto de cambio climático, los modelos tienden a pronosticar con mayor frecuencia condiciones cálidas que frías, las cuales usualmente se verifican.

Tabla 7: Ídem Tabla 4 pero para la temperatura.

Áreas ROC	Modelos	Categorías		
		Superior a lo normal	Normal	Inferior a lo normal
Temperatura	Consenso	0.59	0.56	0.53
	SMN-CPT	0.53	0.61	0.51
	SMN-CPT.NMME	0.55	0.62	0.53
	CLIMAX	0.56	0.55	0.50
	IRI	0.52	0.50	0.49
	NMME	0.57	0.54	0.50

8. DISCUSIÓN Y CONCLUSIONES

En este trabajo se presenta el procedimiento y las metodologías utilizadas para la verificación no sólo del pronóstico estacional por consenso elaborado por el SMN, sino también de las previsiones de otros centros o modelos climáticos, para ser usados con fines comparativos. Los modelos considerados para la comparación son: el NMME, el IRI, el CLIMAX, el SMN-CPT y el SMN-CPT.NMME. El período de análisis abarca desde el trimestre enero-febrero-marzo de 2018 hasta la fecha.

Los resultados de las métricas calculadas permiten concluir que, entre todos los modelos considerados, el consenso es el que presenta el mayor porcentaje de aciertos en los pronósticos más probables. Esto se advierte tanto en la temperatura como en la precipitación. Se destaca el IRI con muy baja frecuencia de observaciones en las predicciones menos probables. Cabe destacar que esta última información no era utilizada para los profesionales encargados de la elaboración del pronóstico por consenso, debido a que la Universidad de Columbia sólo ofrece al público las figuras de las predicciones con probabilidad dominante. Desde julio de 2021 se empezó a construir los campos de los pronósticos menos probables para brindárselos a los pronosticadores. Su construcción implica la descarga y manipulación de los datos, incluyendo el empleo de máscaras.

También se pudo concluir que, para la precipitación, el IRI, el consenso y el CLIMAX tienen buena habilidad en discriminar las condiciones húmedas y secas. Los dos primeros, sin embargo, no logran distinguir las precipitaciones dentro del rango normal. En cuanto a la temperatura, el consenso presenta la mejor performance para discriminar los eventos cálidos y fríos. El SMN-CPT.NMME y el SMN-CPT muestran el máximo desempeño para discriminar las condiciones normales de las no normales, tanto de precipitación como de temperatura. Estos dos modelos, junto con el CLIMAX, exhiben valores extremos de *IS*, debido a que atribuyeron probabilidades mínimas y cercanas a cero a eventos finalmente observados. El consenso, al ser un pronóstico subjetivo, no tiene estos problemas, dado que no asigna probabilidades tan extremas.

Para la verificación y el cálculo de las métricas se tuvo en cuenta el porcentaje de pronósticos informativos o no-climatológicos de cada modelo. El consenso presenta el máximo porcentaje en la mayoría de los trimestres, seguido por el CLIMAX, mientras que el IRI y el NMME, el mínimo.

Este trabajo se enmarca dentro de los esfuerzos del SMN para establecer verificaciones objetivas de todos sus pronósticos. Utiliza como punto de partida las evaluaciones elaboradas por los profesionales responsables del pronóstico trimestral oficial pero incluye un análisis más exhaustivo, incorporando métricas y procedimientos recomendados por la OMM. Si bien este estudio permitió obtener resultados y formular conclusiones importantes, resta explorar otras métricas que no sólo posibiliten validar de manera correcta las predicciones climáticas, sino que sean de fácil comunicación y que puedan ser utilizadas por otros grupos de verificación de la institución. También resta incluir más sistemas de pronósticos

estacionales objetivos para la comparación de desempeños. Como el pronóstico trimestral tiene sus características propias, encontrar métricas comunes representa un gran desafío. La comunicabilidad es otro factor clave, dado que es política del SMN disponer de métricas que sean de fácil comunicación. Esto favorecerá la transmisión de las incertidumbres de las previsiones de largo plazo a los usuarios no especializados.

APÉNDICE A: Casos de estudio

A.1 Mapas de pronóstico individuales

En este apartado se ilustran los pasos para calcular las métricas recomendadas cuando se dispone de un único campo de pronóstico, de un trimestre en particular. Se siguen los lineamientos planteados por Mason (2018), pero empleando un caso de estudio local. Se utiliza un conjunto simple de datos, que corresponde al pronóstico de precipitación acumulada del consenso en el trimestre abril-mayo-junio de 2018 para 22 estaciones meteorológicas. Las estaciones pertenecen a las regiones 1, 2 y 3 de la Figura 7. Éstas se listan en la Tabla A.1, junto el tercil observado y las probabilidades pronosticadas en las categorías “superior a lo normal”, “normal” e “inferior a lo normal”.

Tabla A.1: Ejemplo del pronóstico por consenso de la precipitación para el trimestre abril-mayo-junio de 2018, con el sistema de probabilidades de tres categorías: “superior a lo normal” (“Superior”), “normal” (“Normal”) e “inferior a lo normal” (“Inferior”). En la columna 1 a 3 se encuentran la región, el número y nombre de las estaciones meteorológicas consideradas, respectivamente; en la columna 4, la categoría observada o verificada; y en las columnas 5 a 7, las probabilidades pronosticadas para cada una de las categorías. Las probabilidades de la estación Orán Aero han sido modificadas para fines ilustrativos.

Región	OMM	Nombre	Observación	Superior	Normal	Inferior
1	87097	Iguazú	Inferior	0.25	0.35	0.40
1	87155	Resistencia Aero	Normal	0.25	0.35	0.40
1	87162	Formosa	Inferior	0.25	0.35	0.40
1	87163	Bernardo de Irigoyen Aero	Inferior	0.25	0.35	0.40
1	87166	Corrientes Aero	Inferior	0.25	0.35	0.40
1	87178	Posadas Aero	Inferior	0.25	0.35	0.40
1	87187	Oberá	Inferior	0.25	0.35	0.40
1	87270	Reconquista	Superior	0.25	0.35	0.40
1	87289	Paso de los Libres Aero	Inferior	0.25	0.35	0.40
1	87393	Monte Caseros Aero	Superior	0.25	0.35	0.40
2	87371	Sauce Viejo Aero	Superior	0.25	0.35	0.40
2	87374	Paraná Aero	Superior	0.25	0.35	0.40
2	87395	Concordia Aero	Superior	0.25	0.35	0.40
2	87480	Rosario Aero	Superior	0.25	0.35	0.40
2	87497	Galeguaychú Aero	Superior	0.25	0.35	0.40
2	87484	Pergamino	Superior	0.20	0.45	0.35
2	87494	San Pedro	Superior	0.20	0.45	0.35
3	87148	Pcia. Roque Sáenz	Normal	0.35	0.40	0.25

		Peña				
3	87016	Orán Aero	Normal	0.40	0.40	0.20
3	87022	Tartagal	Normal	0.45	0.35	0.20
3	87078	Las Lomitas	Inferior	0.35	0.40	0.25
3	87129	Santiago del Estero	Superior	0.35	0.40	0.25

Hit scores

El *HS* se calcula a partir de la siguiente ecuación:

$$\bar{y}_k = 100\% \times \frac{1}{n} \sum_{i=1}^n y_{k,i} \quad (\text{A.1})$$

donde n es el número de pronóstico, e $y_{k,i}$ es la cantidad de veces que la observación correspondió a la categoría con la k -ésima mayor probabilidad. Para un sistema de predicción de tres categorías equiprobables, como el que se considera aquí, se recomienda calcular el *HS* no sólo para $k = 1$, la categoría más probable, sino también $k = 2, 3$, es decir para las categorías con segunda y tercera mayor probabilidad.

El *HS* varía del 0% al 100%. Si la categoría con la mayor probabilidad nunca se verifica, su correspondiente *HS* será del 0%, mientras que si siempre se corrobora, será del 100%. En el primer caso se trata de un pronóstico completamente malo, mientras que el segundo, de un pronóstico perfecto; en ambas situaciones, la resolución es fuerte. Los pronósticos sólo serán considerados “buenos” si su *HS* excede el 33% que le asigna la climatología.

En la Tabla A.2 se presenta el procedimiento del cálculo de los *HS* para los datos de la Tabla A.1.

Tabla A.2: Ejemplo del procedimiento del cálculo de los *HS* para los datos de la Tabla A.1. En la columna 1 a 3 se encuentran la región, el número y nombre de las estaciones meteorológicas consideradas, respectivamente; en la columna 4, la categoría observada o verificada; en las columnas 5 a 7, el rango de las categorías pronosticadas en función de sus probabilidades; y en las columnas 8 a 10, los elementos definidos en la Ecuación A.1. Las probabilidades de la estación Orán Aero han sido modificadas para fines ilustrativos.

Región	OMM	Nombre	Observación	Rango de categoría de pronóstico			Aciertos		
				Superior	Normal	Inferior	$y_{1,i}$	$y_{2,i}$	$y_{3,i}$
1	87097	Iguazú	Inferior	3	2	1	1	0	0
1	87155	Resistencia Aero	Normal	3	2	1	0	1	0
1	87162	Formosa	Inferior	3	2	1	1	0	0
1	87163	Bernardo De Irigoyen Aero	Inferior	3	2	1	1	0	0
1	87166	Corrientes Aero	Inferior	3	2	1	1	0	0
1	87178	Posadas Aero	Inferior	3	2	1	1	0	0
1	87187	Oberá	Inferior	3	2	1	1	0	0

1	87270	Reconquista	Superior	3	2	1	0	0	1
1	87289	Paso De Los Libres Aero	Inferior	3	2	1	1	0	0
1	87393	Monte Caseros Aero	Superior	3	2	1	0	0	1
2	87371	Sauce Viejo Aero	Superior	3	2	1	0	0	1
2	87374	Paraná Aero	Superior	3	2	1	0	0	1
2	87395	Concordia Aero	Superior	3	2	1	0	0	1
2	87480	Rosario Aero	Superior	3	2	1	0	0	1
2	87497	Gualeguaychú Aero	Superior	3	2	1	0	0	1
2	87484	Pergamino	Superior	3	1	2	0	0	1
2	87494	San Pedro	Superior	3	1	2	0	0	1
3	87148	Pcia. Roque Sáenz Peña	Normal	2	1	3	1	0	0
3	87016	Orán Aero	Normal	1	1	3	1	0	0
3	87022	Tartagal	Normal	1	2	3	0	1	0
3	87078	Las Lomitas	Inferior	2	1	3	0	0	1
3	87129	Santiago Del Estero	Superior	2	1	3	0	1	0
<i>Número de hits</i> <i>Número de pronósticos</i>							$\frac{9}{22} = 0.41$	$\frac{3}{22} = 0.14$	$\frac{10}{22} = 0.45$

A partir de la Tabla A.2 se observa que el valor del *HS* para la categoría más probables es del 41%, mientras que para las categorías con segunda mayor y menor probabilidad es del 14% y 45%, respectivamente.

Ignorance score

El *IS* se calcula a partir del logaritmo en base 2 de la probabilidad de pronóstico en la categoría verificada, mediante la ecuación:

$$Ignorance\ score = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{j,i} \log_2(p_{j,i}) \quad (A.2)$$

donde n es la cantidad total de estaciones y m es el número total de categorías. En este caso, $m = 3$. Por su parte, $y_{j,i}$ vale 1 si la i -ésima observación corresponde a la categoría j ó 0 de lo contrario, y $p_{j,i}$ es la correspondiente probabilidad del pronóstico en dicha categoría.

El rango de la métrica es de 0 para pronósticos perfectos hasta infinito para pronósticos enteramente malos. Más aún, si un pronóstico asigna una probabilidad de 0% a una categoría luego observada, el valor del *IS* será infinito, aún si todos los demás pronósticos tienen un 100% de probabilidad en la categoría verificada.

En la Tabla A.3 se muestra el procedimiento del cálculo del *IS* para los datos de la Tabla A.1.

Tabla A.3: Ejemplo del procedimiento del cálculo del *IS* para los datos de la Tabla A.1. En la columna 1 a 3 se encuentran la región, el número y nombre de las estaciones meteorológicas consideradas, respectivamente; en la columna 4, la categoría observada o verificada; y en las columnas 5 a 8, los elementos definidos en la Ecuación A.2. Las probabilidades de la estación Orán Aero han sido modificadas para fines ilustrativos.

Región	OMM	Nombre	Observación	i	j	$p_{j,i}$	$\sum_{j=1}^m y_{j,i} \log_2(p_{j,i})$
1	87097	Iguazú	Inferior	1	3	0.40	-1.32
1	87155	Resistencia Aero	Normal	2	2	0.35	-1.51
1	87162	Formosa	Inferior	3	3	0.40	-1.32
1	87163	Bernardo de Irigoyen Aero	Inferior	4	3	0.40	-1.32
1	87166	Corrientes Aero	Inferior	5	3	0.40	-1.32
1	87178	Posadas Aero	Inferior	6	3	0.40	-1.32
1	87187	Oberá	Inferior	7	3	0.40	-1.32
1	87270	Reconquista	Superior	8	1	0.25	-2.00
1	87289	Paso De Los Libres Aero	Inferior	9	3	0.40	-1.32
1	87393	Monte Caseros Aero	Superior	10	1	0.25	-2.00
2	87371	Sauce Viejo Aero	Superior	11	1	0.25	-2.00
2	87374	Parana Aero	Superior	12	1	0.25	-2.00
2	87395	Concordia Aero	Superior	13	1	0.25	-2.00
2	87480	Rosario Aero	Superior	14	1	0.25	-2.00
2	87497	Gualedguaychú Aero	Superior	15	1	0.25	-2.00
2	87484	Pergamino	Superior	16	1	0.20	-2.32
2	87494	San Pedro	Superior	17	1	0.20	-2.32
3	87148	Pcia. Roque Sáenz Peña	Normal	18	2	0.40	-1.32
3	87016	Orán Aero	Normal	19	2	0.40	-1.32
3	87022	Tartagal	Normal	20	2	0.35	-1.51
3	87078	Las Lomitas	Inferior	21	3	0.25	-2.00
3	87129	Santiago del Estero	Superior	22	1	0.35	-1.51
$-\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m y_{j,i} \log_2(p_{j,i}) = 1.69$							

De la Tabla A.3 se advierte que el valor del *IS* es del 1.69.

Average interest rate

El *AIR* se calcula a partir de los cocientes entre la probabilidad de pronóstico en la categoría verificada y la probabilidad climatológica, utilizando la expresión:

$$\text{Average interest rate} = 100\% \times \left[\frac{1}{n} \sum_{i=1}^n \frac{p_i}{c_i} - 1 \right] \quad (\text{A.3})$$

donde p_i es la probabilidad del pronóstico para la categoría verificada en la i -ésima estación de las n totales, y c_i corresponde a la probabilidad climatológica. Aquí $c_i = \frac{1}{3} \approx 0.33$.

Como se expresó anteriormente, el *AIR* es positivo para pronósticos “buenos”. Para un sistema de predicción de tres categorías equiprobables, su rango de valores va desde los -100% hasta los 200%. Se puede entender en términos de ganancias y pérdidas en finanzas: para los pronósticos perfectamente buenos, en donde el 100% del capital ha sido colocado en la categoría finalmente verificada, se obtendrían ganancias del doble de la cantidad invertida y, por lo tanto, un *AIR* del 200%; para los pronósticos perfectamente malos, en el que el 0% del capital ha sido invertido en la categoría corroborada, se generarían pérdidas totales, con un *AIR* de -100%.

En la Tabla A.4 se muestra el procedimiento del cálculo del *AIR* para los datos de la Tabla A.1.

Tabla A.4: Ejemplo del procedimiento del cálculo del *AIR* para los datos de la Tabla A.1. En la columna 1 a 3 se encuentran la región, el número y nombre de las estaciones meteorológicas consideradas, respectivamente; en la columna 4, la categoría observada o verificada; y en las columnas 5 a 8, los elementos definidos en la Ecuación A.3. Las probabilidades de la estación Orán Aero han sido modificadas para fines ilustrativos.

Región	OMM	Nombre	Observación	i	j	$p_{j,i}$	$\frac{p_i}{c_i}$
1	87097	Iguazú	Inferior	1	3	0.40	1.20
1	87155	Resistencia Aero	Normal	2	2	0.35	1.05
1	87162	Formosa	Inferior	3	3	0.40	1.20
1	87163	Bernardo de Irigoyen Aero	Inferior	4	3	0.40	1.20
1	87166	Corrientes Aero	Inferior	5	3	0.40	1.20
1	87178	Posadas Aero	Inferior	6	3	0.40	1.20
1	87187	Oberá	Inferior	7	3	0.40	1.20
1	87270	Reconquista	Superior	8	1	0.25	0.75
1	87289	Paso de los Libres Aero	Inferior	9	3	0.40	1.20
1	87393	Monte Caseros Aero	Superior	10	1	0.25	0.75
2	87371	Sauce Viejo Aero	Superior	11	1	0.25	0.75
2	87374	Parana Aero	Superior	12	1	0.25	0.75
2	87395	Concordia Aero	Superior	13	1	0.25	0.75

2	87480	Rosario Aero	Superior	14	1	0.25	0.75
2	87497	Gualeguaychú Aero	Superior	15	1	0.25	0.75
2	87484	Pergamino	Superior	16	1	0.20	0.60
2	87494	San Pedro	Superior	17	1	0.20	0.60
3	87148	Pcia. Roque Sáenz Peña	Normal	18	2	0.40	1.20
3	87016	Orán Aero	Normal	19	2	0.40	1.20
3	87022	Tartagal	Normal	20	2	0.35	1.05
3	87078	Las Lomitas	Inferior	21	3	0.25	0.75
3	87129	Santiago del Estero	Superior	22	1	0.35	1.05
$\left(\frac{1}{n} \sum_{i=1}^n \frac{p_i}{c_i} \right) - 1 \approx -0.04$							

A partir de la Tabla A.4 se observa que el valor del AIR es de aproximadamente el -4%. Esto indica el porcentaje de pérdida de ganancias que se obtendría al apostar por el pronóstico por consenso en lugar de la climatología.

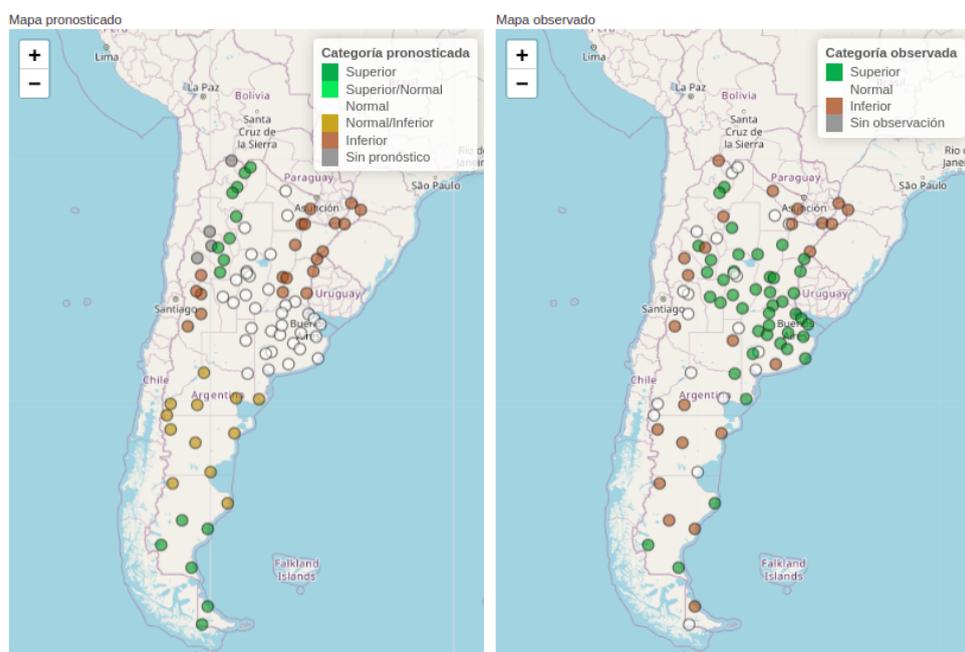


Fig. A.1: Ejemplo del campo pronosticado por el consenso para la precipitación del trimestre abril-mayo-junio de 2018 (izq.) y el campo observado (der.). La información se presenta en función de las categorías “superior a lo normal” (“Superior”), “normal” (“Normal”) e “inferior a lo normal” (“Inferior”). En el mapa del consenso se encuentra graficada la categoría con mayor probabilidad de pronóstico; pueden haber dos categorías con iguales probabilidades (“Superior/Normal” y “Normal/Inferior”).

En la Figura A.1 se muestra el campo pronosticado por el consenso para la precipitación de abril-mayo-junio de 2018, junto con el campo finalmente observado. Las observaciones también se presentan en categorías, de manera de que haya una correspondencia con la información pronosticada. Esto permite no sólo realizar una comparación visual de los mapas, sino también establecer cuál fue el comportamiento de los valores registrados en función de la distribución histórica de la variable.

A.2 Series de pronósticos

Para ilustrar los pasos a seguir para calcular las métricas recomendadas cuando se dispone de una serie de pronóstico, se utiliza el conjunto de datos del ejemplo planteado en Mason (2018). La información del tercil observado y de las probabilidades pronosticadas en las categorías “superior a lo normal”, “normal” e “inferior a lo normal” de este conjunto se presenta en la Tabla A.5. No se consideraron los datos del consenso porque, desafortunadamente, sólo se cuenta con pocos años de pronóstico. En este caso, no es posible calcular métricas o construir diagramas significativos sin tener que recurrir a las técnicas de remuestreo. Hasta que no se disponga de un gran historial de predicciones, los procedimientos deberán realizarse combinando los pronósticos de todas las estaciones del año y ubicaciones.

Tabla A.5: Ejemplo de un conjunto de predicciones y observaciones para realizar los procedimientos de verificación para una serie de pronósticos. La información se presenta siguiendo el sistema de probabilidades de tres categorías: “superior a lo normal” (“Superior”), “normal” (“Normal”) e “inferior a lo normal” (“Inferior”). En la columna 1 se encuentran los años; en la columna 2, la categoría observada o verificada; y en las columnas 3 a 5, las probabilidades pronosticadas de cada una de las categorías. Corresponde a la Tabla B.1 de Mason (2018).

Año	Observación	Inferior	Normal	Superior
2001	Inferior	0.45	0.35	0.20
2002	Inferior	0.50	0.30	0.20
2003	Inferior	0.35	0.40	0.25
2004	Inferior	0.33	0.33	0.33
2005	Normal	0.25	0.35	0.40
2006	Normal	0.20	0.35	0.45
2007	Superior	0.20	0.35	0.45
2008	Superior	0.25	0.40	0.35

Relative operating characteristics

Para medir la discriminación del pronóstico para cada una de las categorías, se recomienda utilizar como métrica el *área ROC*, es decir, el área debajo de la *curva ROC*. El área A se puede calcular a partir de la siguiente expresión:

$$A = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} I(p_{0,i}, p_{1,j}) \quad (\text{A.4})$$

donde n_1 es el número de eventos (es decir, la cantidad de veces que ocurrió la categoría pronosticada en cuestión), n_0 es el número de no eventos (el cantidad de veces que ésta no sucedió), $p_{1,j}$ es la probabilidad

pronosticada para el j -ésimo evento observado, $p_{0,j}$ es la probabilidad de un evento para el i -ésimo no evento, e $I(p_{0,i}, p_{1,i})$ se define como:

$$\begin{aligned}
 I(p_{0,i}, p_{1,i}) &= 0 \text{ si } p_{1,j} < p_{0,i} \\
 I(p_{0,i}, p_{1,i}) &= 0.5 \text{ si } p_{1,j} = p_{0,i} \\
 I(p_{0,i}, p_{1,i}) &= 1 \text{ si } p_{1,j} > p_{0,i}
 \end{aligned}
 \tag{A.5}$$

En la Tabla A.6 se muestra el procedimiento del cálculo del área ROC para los datos de la Tabla A.5, estableciendo la categoría “superior a lo normal” como evento y el resto de las categorías, como no eventos. Como pasos a seguir, se debe considerar todos los años cuyas observaciones fueron inferiores a las normales o normales, y se los debe comparar con cada uno de los años en donde los registros fueron superiores a los normales. Si la probabilidad de la categoría “superior a lo normal” es mayor en el año que efectivamente ocurrió dicha categoría, los años se discriminaron con éxito y la puntuación obtenida será de 1. En caso contrario, la puntuación será de 0 ó 0.5. Luego, se debe hacer una sumatoria sobre todas las puntuaciones.

Tabla A.6: Ejemplo del procedimiento del cálculo del área ROC para los datos de la Tabla A.5 y estableciendo la categoría “superior a lo normal” como evento. En las columnas se encuentran los elementos definidos en las Ecuaciones A.4 y A.5. Corresponde a la Tabla B.3 de Mason (2018).

i	j	Año _{0,i}	Año _{1,j}	$p_{0,i}$	$p_{1,j}$	$I(p_{0,i}, p_{1,i})$
1	1	2001	2007	0.20	0.45	1.0
1	2	2001	2008	0.20	0.35	1.0
2	1	2002	2007	0.20	0.45	1.0
2	2	2002	2008	0.20	0.35	1.0
3	1	2003	2007	0.25	0.45	1.0
3	2	2003	2008	0.25	0.35	1.0
4	1	2004	2007	0.33	0.45	1.0
4	2	2004	2008	0.33	0.35	1.0
5	1	2005	2007	0.40	0.45	1.0
5	2	2005	2008	0.40	0.35	0.0
6	1	2006	2007	0.45	0.45	0.5
6	2	2006	2008	0.45	0.35	0.0
$\sum I(p_{0,i}, p_{1,i}) = 9.5$						

Como en este caso $n_1 = 2$ y $n_0 = 6$, a partir de los resultados de la Tabla A.6 se obtiene que $A = \frac{1}{12} \times 9.5 = 0.79$.

Para cada categoría o evento, la *curva ROC* se construye calculando los *HR* y los *POFD* para umbrales de probabilidad decrecientes. Los *HR* y *POFD* se calculan usando las siguientes expresiones:

$$y_k = 100 \times \left(\frac{1}{n_1} \sum_{i=1}^{n_1} y_{k,i} \right) \quad (\text{A.6})$$

$$x_k = 100 \times \left(\frac{1}{n_0} \sum_{i=1}^{n_0} y_{k,i} \right) \quad (\text{A.7})$$

donde n_1 es el número de eventos, n_0 es el número de no eventos, e y_k y x_k es el número de veces que ocurrió un evento o un no evento dada una probabilidad de pronóstico que excede el k -ésimo umbral de probabilidad, respectivamente.

En la Tabla A.7 se presenta el procedimiento para el cálculo de la *curva ROC* considerando, nuevamente, los datos de la Tabla A.5 y estableciendo la categoría “superior a lo normal” como el evento.

Tabla A.7: Ejemplo del procedimiento de la construcción de la *curva ROC* para los datos de la Tabla A.5 y estableciendo la categoría “superior a lo normal” como evento. En las columnas se encuentran los elementos definidos en las Ecuaciones A.6 y A.7. Se contemplan seis umbrales de probabilidad (0.45, 0.40, 0.35, 0.33, 0.30, 0.25, 0.20). Corresponde a la Tabla B.4.a de Mason (2018).

Año	Evento	p	Umbrales de probabilidad						
			0.45	0.40	0.35	0.33	0.30	0.25	0.20
2001	No	0.20	0	0	0	0	0	0	1
2002	No	0.20	0	0	0	0	0	0	1
2003	No	0.25	0	0	0	0	0	1	1
2004	No	0.33	0	0	0	1	1	1	1
2005	No	0.40	0	1	1	1	1	1	1
2006	No	0.45	1	1	1	1	1	1	1
<i>FAR</i>			0.17	0.33	0.33	0.50	0.50	0.67	1.00
2007	Yes	0.45	1	1	1	1	1	1	1
2008	Yes	0.35	0	0	1	1	1	1	1
<i>HR</i>			0.5	0.5	1	1	1	1	1

En la Tabla A.7 se ve que, para todos los años con o sin evento, se indica con un 1 si la probabilidad de pronóstico por encima de lo normal es mayor o igual que un determinado umbral de probabilidad, y con un 0 en caso contrario. El gráfico se construye trazando los *HR* en el eje horizontal versus los *POFD* en el eje vertical y se presenta en la Figura A.2.

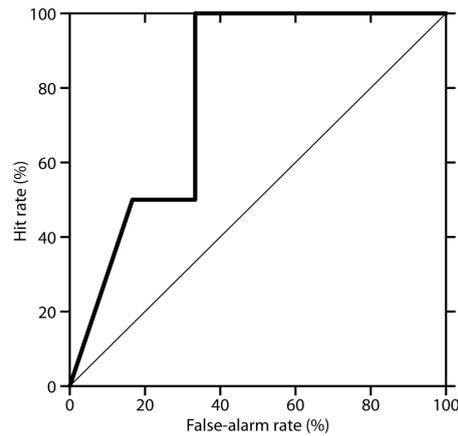


Fig. A.2: Ejemplo de curva ROC para los datos de la Tabla A.5 y estableciendo la categoría “superior a lo normal” como evento. Corresponde a la Figura B.1 de Mason (2018).

Diagrama de confiabilidad

Para cada categoría, el *diagrama de confiabilidad* traza la frecuencia observada versus la probabilidad de pronóstico, en donde el rango de probabilidades del pronóstico se divide en k intervalos (por ejemplo, en los intervalos 0-5%, 5-15%, 15-25%, etc.). A partir de dicho diagrama se determina qué tan bien se corresponden las probabilidades predichas de un evento con sus frecuencias observadas.

El procedimiento para la construcción de los diagramas de confiabilidad no se presenta, dado que sus resultados no serían robustos en función de la información de la Tabla A.5. Este requiere de un conjunto de datos relativamente grande, debido a la necesidad de dividir y agrupar la muestra en submuestras condicionadas a la probabilidad de pronóstico.

APÉNDICE B

Intervalos de confianza

Cuando se obtiene una estimación puntual de un parámetro (en este caso, la estimación directa de una dada métrica de verificación), es recomendable acompañar dicha estimación con una medida de su precisión, como es el intervalo de confianza. El intervalo de confianza está definido como un rango de valores que, con una probabilidad determinada por el nivel de confianza, contiene al verdadero valor del parámetro en cuestión.

La técnica de bootstrap o remuestreo puede ser utilizada cuando los intervalos no pueden definirse de manera analítica. El bootstrap permite recrear el proceso de extracción de muestras de una población pero, en lugar de obtener repetidamente conjuntos de datos de dicha población, se lo hace a partir de la muestra original.

Mason (2018) describe el procedimiento de bootstrap con reposición para definir el intervalo de confianza de una métrica de verificación de la siguiente manera: si se tiene una muestra original con n pares de datos de pronóstico-observación, para obtener o extraer una nueva muestra (una muestra de bootstrap) a partir de ella, se deben seleccionar n pares de datos con reemplazo o reposición. La reposición de elementos va a determinar que ciertos pares de pronóstico-observación sean seleccionados más de una vez y otros, nunca. Luego, se calcula la métrica de verificación para esta muestra de bootstrap. Estos pasos deben repetirse nb veces, obteniéndose así nb muestras de bootstrap y nb métricas de verificación asociadas. Se recomienda que nb sea, al menos, 1000. Finalmente, las nb métricas de verificación son utilizadas para estimar el intervalo de confianza. Para esto se pueden utilizar diferentes métodos detallados más abajo.

James y otros (2013) ilustran y ejemplifican el proceso de remuestreo con reposición a partir de su Figura 5.11 (Figura B.1).

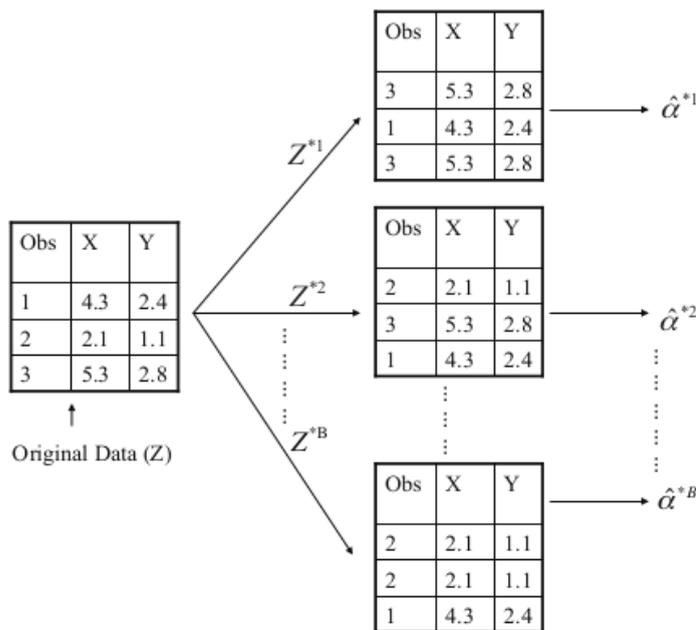


Fig. B.1: Ilustración gráfica del enfoque bootstrap en una pequeña muestra que contiene $n = 3$ observaciones. Cada conjunto de datos bootstrap contiene n observaciones, muestreadas con reemplazo del conjunto de datos original, y se utiliza para obtener una estimación de α . Corresponde a la Figura 5.11 de James y otros (2013).

A partir de la Figura B.1 se ve que, inicialmente, se considera una muestra pequeña Z de tan sólo $n = 3$ pares de datos, en donde los pares observación-pronóstico están representados por los pares X - Y . De esta muestra original, de la cual se obtuvo previamente una estimación $\hat{\alpha}$ del parámetro α , se seleccionan tres elementos aleatoriamente, a fin de producir la primera muestra de bootstrap Z^{*1} . La muestra Z^{*1} contiene al tercer par de datos dos veces y al primero, una sola vez; sin embargo, no incluye al segundo par. Luego, se utiliza a Z^{*1} para calcular una estimación $\hat{\alpha}^{*1}$. Esto lo realizan $nb = B$ veces, obteniendo así B muestras de bootstrap ($Z^{*1}, Z^{*2}, \dots, Z^{*B}$) con sus B estimaciones de α correspondientes ($\hat{\alpha}^{*1}, \hat{\alpha}^{*2}, \dots, \hat{\alpha}^{*B}$).

Para la construcción de los intervalos de confianza, Mason (2018) propone utilizar el método de percentiles, el cual también se detalla en Efron y Gong (1983). Este y otros métodos para tal fin están descritos en Efron y Tibshirani (1993). Para este apéndice también se consideró como referencia la tesis de Kerner (2015), utilizando la notación planteada por James y otros (2013) del ejemplo anterior.

El método más simple para abordar la construcción de los intervalos de confianza es el **método normal**.

Este supone que, nuevamente, si $\hat{\alpha}$ es una estimación del valor real de α y $\hat{\alpha}^{*1}, \hat{\alpha}^{*2}, \dots, \hat{\alpha}^{*B}$ son las estimaciones bootstrap (también llamadas replicaciones o réplicas bootstrap), dichas estimaciones tienen una distribución asintóticamente normal. Luego, se define el intervalo de confianza bootstrap normal para α con un nivel de confianza del 90% de la siguiente manera:

$$IC_{normal} = [\hat{\alpha} - 1.645 \cdot \widehat{se}_B; \hat{\alpha} + 1.645 \cdot \widehat{se}_B] \quad (B.1)$$

donde \widehat{se}_B es el desvío estándar de las estimaciones bootstrap y se calcula utilizando la siguiente ecuación:

$$\widehat{se}_B = \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\alpha}^{*b})^2 - \left(\frac{1}{B} \sum_{b=1}^B \hat{\alpha}^{*b}\right)^2} \quad (B.2)$$

Cabe destacar que la utilización del intervalo definido por la Ecuación B.1 carece de sentido si no se satisface la suposición de normalidad de la distribución.

A diferencia del método normal, el intervalo de confianza construido a través del **método de los percentiles** respeta la distribución empírica de las replicaciones bootstrap, no presuponiendo su normalidad. Este método infiere los límites del intervalos de confianza directamente utilizando los valores de los percentiles. Considerando un nivel de confianza del 90%, el intervalo de confianza bootstrap percentil está definido por la siguiente expresión:

$$IC_{perc} = [\hat{\alpha}^{*(0.05)}, \hat{\alpha}^{*(0.95)}] \quad (B.3)$$

donde $\hat{\alpha}^{*(0.05)}$ y $\hat{\alpha}^{*(0.95)}$ son los percentiles 5 y 95, respectivamente, de la distribución de las réplicas bootstrap. Si bien la Ecuación B.3 considera un escenario ideal, en donde B (el número de repeticiones del algoritmos bootstrap) es infinito, en la práctica éste es un número finito, aunque sí debe ser lo suficientemente grande para poder aproximar con precisión los percentiles altos y bajos de una distribución. Por ello, se utiliza una expresión aproximada de dicha ecuación, que emplea los percentiles empíricos de las réplicas bootstrap. Por ejemplo, si $B = 2000$ y el nivel de confianza es del 90%, el intervalo de confianza bootstrap percentil aproximado se extiende desde el valor 100-ésimo hasta el 1900-ésimo de la lista de las 2000 estimaciones bootstrap ordenadas de manera creciente.

Cabe destacar que si la distribución de las réplicas bootstrap es aproximadamente normal, los intervalos de confianza generados por ambos métodos son similares. Contrariamente al método normal, el intervalo de confianza construido a partir del método del percentil sí preserva el rango de valores posibles del parámetro de interés. Esto es especialmente relevante cuando el parámetro toma sólo un cierto rango de valores, como es el caso del *HS*. Sin embargo, este método es sensible a los outliers y, en caso de haberlos, no corrige sesgos, entre otros defectos.

El **método BCa** (del inglés, *Bias corrected and Accelerated*) es una versión mejorada del método de los percentiles, tanto en la teoría como en la práctica. Fue propuesto por Efron (1987) con el objetivo de corregir ciertas deficiencias de los anteriores métodos. Su construcción depende de dos parámetros: z_0 y a . El primer parámetro cumple la función de corrector de sesgo mientras que el segundo, denominado constante de aceleración, hace referencia a la tasa de cambio del desvío estándar de $\hat{\alpha}$ respecto al verdadero valor α (Di Ciccio y Efron, 1996). Estos parámetros cobran relevancia en los casos de estadísticos $\hat{\alpha}$ sesgados y con varianza variable. Dado que el método BCa produce los intervalos de confianza más precisos, es recomendable su uso en la mayoría de los casos, principalmente para los problemas no paramétricos. Sin embargo, hay que contemplar su complejidad matemática, que puede no ser beneficiosa para ciertos propósitos. La expresión detallada del intervalo de confianza bootstrap BCa se encuentra en el capítulo 14 y 22 de Efron y Tibshirani (1993).

BIBLIOGRAFÍA

- Barnston, A.G., 1994: Linear Statistical Short-Term Climate Predictive Skill in the Northern Hemisphere. *J. Climate*, 7, 1513–1564, [https://doi.org/10.1175/1520-0442\(1994\)007<1513:LSSTCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<1513:LSSTCP>2.0.CO;2).
- Barnston, A.G., Li, S., Mason, S.J., DeWitt, D.G., Goddard, L. and Gong, X., 2010: Verification of the First 11 Years of IRI's Seasonal Climate Forecasts. *J. Appl. Meteor. Climatol.*, 49, 493–520.
- de Elía, R., 2021: Plan transversal de verificación de pronósticos del SMN. Nota Técnica SMN 2021-84. <http://repositorio.smn.gob.ar/handle/20.500.12160/1435>.
- DiCiccio, T.J. and Efron, B., 1996: Bootstrap confidence intervals. *Statist. Sci.* 11 (3) 189 - 228. <https://doi.org/10.1214/ss/1032280214>.
- Efron, B., 1979: Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7, 1–26. <https://doi.org/10.1214/aos/1176344552>.
- Efron, B., 1987: Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, 82, 171-185. <http://dx.doi.org/10.1080/01621459.1987.10478410>.
- Efron, B. and Gong, G., 1983: A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, 37(1), 36–48. <https://doi.org/10.2307/2685844>.
- Efron, B. and Tibshirani, R.J., 1993: *An Introduction to the Bootstrap*. Chapman & Hall. New York.
- Hagedorn, R. and Smith, L.A., 2008: Communicating the value of probabilistic forecasts with weather roulette. *Meteorological Applications*, 16, <https://doi.org/10.1002/met.92>.
- Hartmann, H.C., Pagano, T.C., Sorooshian, S. and Bales, R., 2002: Confidence builder: Evaluating seasonal climate forecasts from user perspectives. *Bull Amer. Met. Soc.*, 84, 683-698.
- Harvey, L.O., Hammond, K.R., Lusk, C.M. and Mross, E.F., 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.*, 120, 863–883.
- Hogan, R.J. and Mason, I.B., 2012: Deterministic forecasts of binary events. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Chichester, Wiley:31–59, <https://doi.org/10.1002/9781119960003.ch3>.
- James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013: *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Jin, E.K., Kinter, J.L., Wang, B., and Coauthors, 2008: Current status of ENSO prediction skill in coupled ocean–atmosphere models. *Clim. Dyn.*, 31, 647–664, <https://doi.org/10.1007/s00382-008-0397-3>.
- Kerner, 2015: *Análisis Estadístico con el Método Bootstrap: Aplicaciones en Problemas de Regresión*. Tesis de licenciatura. Universidad de Buenos Aires. http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2015/Gaspard_Kerner.pdf
- Kirtman, B.P. and Coauthor, 2014: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bull. Amer. Meteor. Soc.*, 95, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Mason, S.J., 2012: Seasonal and longer-range forecasts. In: *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (I.T. Jolliffe and D.B. Stephenson, eds.). Chichester, Wiley, 31–59, <https://doi.org/10.1002/9781119960003.ch11>.

- Mason, S.J., 2018: Guidance on Verification of Operational Seasonal Climate Forecasts. World Meteorological Organization, Commission for Climatology XIV Technical Report.
- Mason, S.J., Tippet, M.K., Song, L., and Muñoz, A.G., 2020: Climate Predictability Tool version 16.5.8. Columbia University Academic Commons. <https://doi.org/10.7916/d8-em5q-0f07>.
- NCAR - Research Applications Laboratory, 2015: Verification: Weather Forecast Verification Utilities. R package version 1.42. <https://CRAN.R-project.org/package=verification>
- Osman, M., Coelho, C.A.S. and Vera, C.S., 2021: Calibration and combination of seasonal precipitation forecasts over South America using Ensemble Regression. Clim. Dyn. <https://doi.org/10.1007/s00382-021-05845-2>.
- Palmer, T.N. and Coauthors, 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull. Amer. Meteor. Soc., 85, 853–872, <https://doi.org/10.1175/BAMS-85-6-853>.
- R Core Team, 2021: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Roulston, M.S. and Smith, L.A., 2002: Evaluating probabilistic forecasts using information theory. Mon. Wea. Rev., 130, 1653–1660.
- Unger, D.A., van den Dool, H., O'Lenic, E. and Collins, D., 2009: Ensemble Regression. Mon. Wea. Rev., 137, 2365–2379, <https://doi.org/10.1175/2008MWR2605.1>.
- Vigaud, N., Robertson, A.W. and Tippet, M.K., 2017: Multimodel Ensembling of Subseasonal Precipitation Forecasts over North America. Mon. Wea. Rev., 145, 3913–3928, <https://doi.org/10.1175/MWR-D-17-0092.1>.
- Wilks, D.S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteorological Applications, 16(3), 361–368, <https://doi.org/10.1002/met.134>.
- WMO, 2000: Guidelines on Performance Assessment of Public Weather Services, TD No. 1023. https://library.wmo.int/index.php?lvl=notice_display&id=11898#.YVtdjJqZOM8

Instrucciones para publicar Notas Técnicas

En el SMN existieron y existen una importante cantidad de publicaciones periódicas dedicadas a informar a usuarios distintos aspectos de las actividades del servicio, en general asociados con observaciones o pronósticos meteorológicos.

Existe no obstante abundante material escrito de carácter técnico que no tiene un vehículo de comunicación adecuado ya que no se acomoda a las publicaciones arriba mencionadas ni es apropiado para revistas científicas. Este material, sin embargo, es fundamental para plasmar las actividades y desarrollos de la institución y que esta dé cuenta de su producción técnica. Es importante que las actividades de la institución puedan ser comprendidas con solo acercarse a sus diferentes publicaciones y la longitud de los documentos no debe ser un limitante.

Los interesados en transformar sus trabajos en Notas Técnicas pueden comunicarse con Ramón de Elía (rdelia@smn.gov.ar), Luciano Vidal (lvidal@smn.gov.ar) o Martín Rugna (mrugna@smn.gov.ar) de la Dirección Nacional de Ciencia e Innovación en Productos y Servicios, para obtener la plantilla WORD que sirve de modelo para la escritura de la Nota Técnica. Una vez armado el documento deben enviarlo en formato PDF a los correos antes mencionados. Antes del envío final los autores deben informarse del número de serie que le corresponde a su trabajo e incluirlo en la portada.

La versión digital de la Nota Técnica quedará publicada en el Repositorio Digital del Servicio Meteorológico Nacional. Cualquier consulta o duda al respecto, comunicarse con Melisa Acevedo (macevedo@smn.gov.ar).