

Verificación de los pronósticos del Sistema de Asimilación y Pronóstico numérico del SMN para el período 2020-2022.

Nota Técnica SMN 2023-145

Cynthia Matsudo¹ y Yanina García Skabar¹

¹Dirección de Productos de Modelación Ambiental y Sensores Remotos, Dirección Nacional de Ciencia e Innovación en Productos y Servicios, Servicio Meteorológico Nacional

Julio 2023

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 de que la fuente sea adecuadamente citada.

Resumen

Esta nota técnica se desarrolla en el marco del Plan de Verificación Transversal de pronóstico del Servicio Meteorológico Nacional (SMN). Aquí se presentan los resultados de la verificación de los pronósticos numéricos del Sistema de Asimilación y Pronóstico numérico del SMN (SAP.SMN) para el periodo 2020-2022.

Se analizan el desempeño de los pronósticos del esquema determinístico (SAP.SMN-DET) y del ensamble (SAP.SMN-ENS) de las siguientes variables: temperatura a 2m, magnitud del viento a 10m, temperatura mínima y máxima diaria y precipitación acumulada en 24 horas. Las observaciones para la verificación provienen de la red de estaciones de superficie del SMN. En la medida de lo posible, se incluye una comparación con los pronósticos de menor resolución provenientes del Global Forecast System (GFS).

En líneas generales, el SAP.SMN muestra un mejor desempeño que el GFS tanto para los valores calibrados como sin calibrar. En cuanto a precipitación acumulada en 24h, el modelo WRF muestra una subestimación media en el volumen total y presenta una mejor correspondencia a la distribución que el GFS tanto la componente determinística como en los miembros del ensamble. Los valores de Brier Score resultan similares a otros sistemas de pronósticos por ensamble regionales y a partir de los valores de confiabilidad se detecta que existe un error sistemático por lo que, a futuro, podría requerir la aplicación de un método de calibración. De acuerdo a las variables analizadas, el SAP.SMN-ENS resulta subdispersivo. La temperatura a 2 m y la magnitud del viento a 10 m calibradas muestran una sobreestimación con errores de aproximadamente de 2°C y 2 m/s respectivamente. Las temperaturas mínimas y máximas diarias calibradas presentan un desempeño superior al 60% y 70 % en cuanto a porcentaje de acierto respectivamente, superior al ensamble de GFS. El impacto de la calibración se ve más claramente en el caso de las temperaturas máximas diarias.

Los resultados obtenidos pueden responder a las características en la configuración del sistema de pronóstico, así como a la calidad de información utilizada como condición inicial y de borde. A futuro, y con la implementación operativa de una componente de asimilación de datos regional se podrá evaluar el impacto de contar con análisis enriquecidos con información de distintas fuentes locales.

Abstract

This technical note is developed within the framework of the National Meteorological Service (NMS) cross-sectional forecast verification plan. Here we present the results of the verification of the numerical forecasts of the NMS Numerical Assimilation and Forecast System (SAP.SMN) for the period 2020-2022.

The verification for the deterministic scheme (SAP.SMN-DET) and the ensemble scheme (SAP.SMN-ENS) is made for the following variables: 2-m temperature, 10-m wind speed, daily minimum and maximum temperature and 24-hour accumulated precipitation. Observations come from the NMS

surface station network. Where possible, a comparison with lower resolution forecasts from the Global Forecast System (GFS) is included.

In general terms, SAP.SMN shows a better performance than GFS for both calibrated and uncalibrated values. In terms of 24-h accumulated precipitation, WRF model shows an underestimation in the total mean volume and presents a better correspondence to the observed distribution than the GFS for both the deterministic scheme and the ensemble members. Brier Score values are similar to other regional ensemble forecasting systems and from the reliability values a systematic error is detected which could require, in the future, the application of a calibration method. According to the variables analyzed, SAP.SMN-ENS is subdispersive. The 2-m calibrated temperature and 10-m wind speed show an overestimation with errors of approximately 2°C and 2 m/s respectively. Calibrated daily minimum and maximum temperatures show a performance above 60% and 70% in terms of success percentage respectively, superior to what GFS ensemble shows. The impact of calibration is more clearly seen in the case of daily maximum temperatures.

These results may respond to the characteristics of the forecast system configuration, as well as to the quality of the information used as initial and boundary conditions. In the future, and with the operational implementation of a regional data assimilation system, it will be possible to evaluate the impact of having analyses enriched with information from different local sources.

Palabras clave: Verificación, pronóstico, ensamble, WRF, GFS

Citar como: Matsudo C., Y. Garcia Skabar, 2023: Verificación de los pronósticos del Sistema de Asimilación y Pronóstico numérico del SMN para el período 2020-2022. Nota Técnica SMN 2023-145.

1. INTRODUCCIÓN

Desde enero de 2020, y junto con la instalación de un sistema de cómputo de alto rendimiento Huayra Muyu (HM), el Servicio Meteorológico Nacional (SMN) comenzó a producir pronósticos de alta resolución para todo el país con el modelo WRF. Luego de múltiples pruebas de configuración y optimización, quedó implementado en forma operativa un sistema de pronóstico denominado “Sistema de Asimilación y Pronóstico numérico del SMN” (SAP.SMN) que cuenta con una componente determinística (SAP.SMN-DET) y otra probabilística (SAP.SMN-ENS). Los detalles de las configuraciones finales del SAP.SMN-DET se pueden encontrar en Dillon y otros (2020a) y las del ensamble de 20 miembros en Dillon y otros (2020b).

Todo sistema de pronóstico numérico debe contar con un esquema de verificación para cuantificar los distintos tipos de errores en los pronósticos, medir su desempeño y evaluar su utilidad. En el caso de los pronósticos determinísticos, suelen utilizarse herramientas estadísticas simples de sesgo y de error cuadrático medio que pueden aplicarse a un único punto o a varios puntos de un área de interés. Los resultados de la verificación de los pronósticos numéricos del SMN del esquema determinístico para el año 2020 pueden encontrarse en la nota técnica SMN 2020-95 (Matsudo y otros, 2021). En el caso de los pronósticos probabilísticos se puede cuantificar su precisión a partir de medir el rango de escenarios posibles y, por ende, la incertidumbre que logran representar. Dicha incertidumbre se asocia a la representación aproximada del modelo numérico de los procesos que interactúan en el sistema climático, a la naturaleza caótica de la atmósfera y la falta de conocimiento de las condiciones iniciales. En los pronósticos por ensambles la incertidumbre está dada por la dispersión del conjunto. Un ensamble de buena calidad es preciso y además confiable. Esto último implica una dependencia sensible a las condiciones iniciales ya que, en un sistema caótico, dada una pequeña perturbación en las condiciones de entrada pueden llevar a cambios significativos en los pronósticos resultantes. En general, las medidas estadísticas que se suelen considerar para medir el desempeño de los pronósticos por ensambles se aplican al valor medio del conjunto (media del ensamble), al valor de un miembro individual o a la distribución de probabilidad del conjunto de miembros (dispersión del conjunto).

Por otra parte, dentro de las diversas técnicas que se utilizan para disminuir los errores sistemáticos de los pronósticos y mejorar su desempeño, se encuentran las técnicas de calibración. Mediante la aplicación de correcciones adecuadas para cada variable se puede lograr que los errores entre los pronósticos y las observaciones se tornen más pequeños. Los detalles de las técnicas de calibración que se aplican en los pronósticos de temperatura a 2 m, magnitud de viento a 10 m y temperatura mínima y máxima diaria de los pronósticos del SAP.SMN se pueden encontrar en Cutraro y otros (2022, 2020).

El objetivo de esta Nota Técnica es presentar una síntesis del desempeño de los pronósticos determinísticos y probabilísticos generados por el modelo WRF para el periodo 2020-2022 de precipitación acumulada en 24h y temperatura a 2 m, magnitud del viento a 10 m, temperatura mínima y máxima diarias. En la medida de lo posible, se incluye en el estudio los pronósticos de

variables calibradas, así como una comparación con los pronósticos del Global Forecast System (GFS) de menor resolución.

2. DATOS Y METODOLOGÍA

Para este trabajo se utilizan los pronósticos operativos del modelo WRF con núcleo dinámico Advanced Research WRF (ARW) versión 4.0 desarrollado por National Center for Atmospheric Research (NCAR de sus siglas en inglés, Skamarock y otros, 2019). El WRF es un modelo de mesoescala de última generación, compresible, no hidrostático que permite un paso de tiempo de integración variable para resolver la microfísica de forma explícita. El dominio abarca todo Argentina con una proyección Conforme de Lambert (1000x1250 puntos, Figura 1), la resolución horizontal es de 4 km, con 45 niveles verticales (tope 10 hPa). Se ejecutan 4 inicializaciones al día, a las 00, 06, 12 y 18 UTC, y las salidas que se obtienen son horarias de hasta un plazo de 72 horas para el esquema determinístico y de hasta 48 horas para el ensamble. Las condiciones iniciales y de borde (CIs/CBs) provienen de los análisis y pronósticos del GFS producido por el National Center of Environmental Prediction (NCEP de sus siglas en inglés).



Figura 1: Dominio WRF con proyección Conforme de Lambert.

El SAP.SMN-ENS consiste en un ensamble multifísico de 20 miembros definidos por la combinación de 3 parametrizaciones de capas límites (Mellor-Yamada-Janjic, MYJ; Universidad de Yonsei, YSU; Shin Hong, SH) y dos parametrizaciones de la microfísica de 1 momento (WSM6 y LIN) (ver Tabla I). El SAP.SMN-DET utiliza las mismas parametrizaciones que el miembro 001 del SAP.SMN-ENS. Para ambos esquemas, las CIs/CBs provienen de los análisis y pronósticos del GFS de resolución 0.25°, perturbados con los miembros del ensamble de GFS de resolución 0.5°, por lo que cada miembro utiliza CIs/CBs distintas.

Tabla I: Combinación de las configuraciones físicas de los miembros del SAP.SMN-ENS.

Combinación configuración de los miembros del ensamble					
Microfísica	Nro de miembro- capa límite				
WSM6	001-MYJ	002-MYJ	003-MYJ	004-SH	005-SH
	006-SH	007-YSU	008-YSU	009-YSU	010-MYJ
LIN	011-MYJ	012-MYJ	013-MYJ	014-SH	015-SH
	016-SH	017-YSU	018-YSU	019-YSU	020-YSU

Las variables pronosticadas que se verifican son: temperatura a 2 metros (T2m), temperatura a 2 metros calibrada (T2cal), temperatura mínima y máxima sin calibrar y calibradas (Tmín_f, Tmáx_f, Tmín_cal, Tmáx_cal respectivamente), magnitud del viento a 10m sin calibrar (magviento) y calibrada (magVientocal) y precipitación acumulada en 24 horas (PP24). Para todos los ciclos, la temperatura mínima corresponde al valor mínimo horario pronosticado entre 00 y 12Z (horas de la mañana) mientras que la temperatura máxima corresponde al valor máximo horario pronosticado entre 12 y 00Z (horas de la tarde). Se verifican los valores correspondientes al “día+1” siendo “día” la fecha de análisis de la corrida. Los respectivos valores calibrados (Tmáx_cal y Tmín_cal) se obtienen aplicando la misma metodología de calibración que para T2cal (Cutraro y otros, 2020). En el caso de las Tmín_cal y Tmáx_cal de GFS solo están disponibles a partir de julio de 2020. Si bien los pronósticos de precipitación tienen resolución horaria, en este informe se verifica la precipitación acumulada de 24 horas, entre las 9 hs de un día y el siguiente, que coincide con el día pluviométrico.

El desempeño de los pronósticos se midió utilizando las observaciones provenientes de la red de estaciones convencionales de superficie del SMN correspondientes al periodo 2020-2022. Por cuestiones metodológicas y de disponibilidad de información, para este informe se consideraron 101 estaciones meteorológicas (ver ANEXO I). Cabe destacar, que el régimen de medición de la red de estaciones no es uniforme por lo que la disponibilidad de observaciones a lo largo del día depende de cada estación.

Para el cálculo de los estadísticos de verificación (ver ANEXO II), los datos observados se comparan con el dato del punto más cercano de la retícula de cada modelo. En este sentido, se contempla que dicho punto no se ubique sobre el agua en el caso de estaciones cercanas a la costa y tengan la mayor correspondencia con la altura de la estación (más detalles en Cutraro y otros, 2020).

3. RESULTADOS Y DISCUSIÓN

A continuación, se muestran los resultados del desempeño del modelo WRF para la componente determinística y probabilística. En la mayor parte de los casos, se presentan los resultados para la inicialización de 06 UTC dado que los otros ciclos muestran resultados similares.

3.1 SAP.SMN-DET

3.1.1 Precipitación

En primer lugar, se hace una inspección del desempeño de la serie anual de PP24 media en todo el país (figura 2). Se puede ver que en los meses de primavera/verano (diciembre a abril) se observa una subestimación siendo mayor para el modelo WRF que para GFS para ambos plazos de 30 y 54h. En particular, la serie de RMSE muestra que para los meses de enero a abril la subestimación para el modelo WRF promedia los 10 mm mientras que en diciembre se mantiene por debajo.

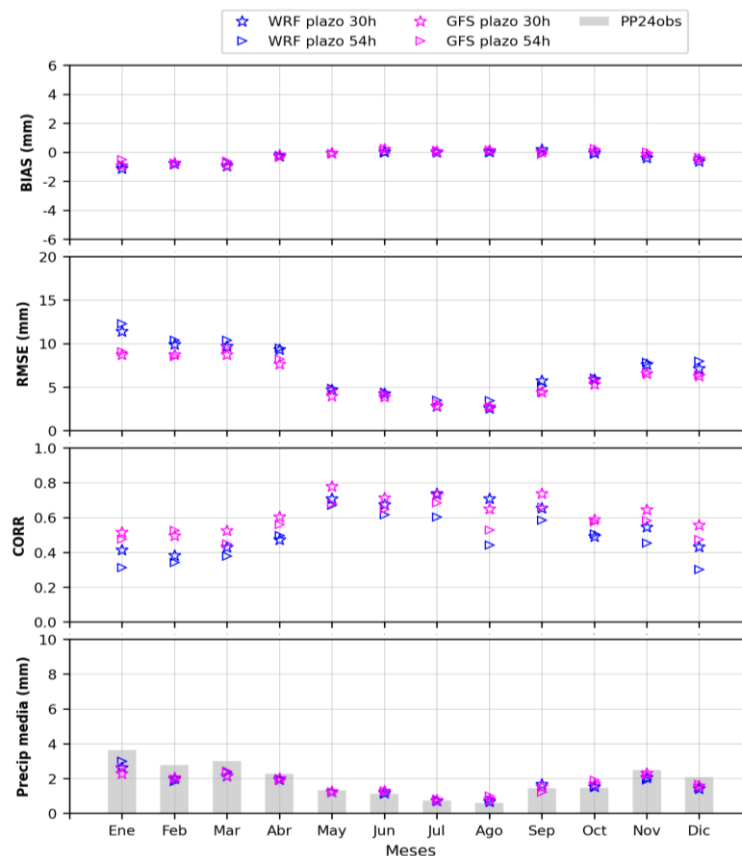


Figura 2: Serie anual de estadísticas continuas de la PP24 por WRF y GFS (ciclo 06Z, plazo 30h). Las barras grises en el último panel indican el valor de PP24 total observado en todas las estaciones para cada mes.

La distribución de los errores se puede ver en la figura 3. El modelo WRF muestra que la mayor subestimación se observa en las estaciones del centro y noreste del país mientras que las estaciones más cercanas a la cordillera, así como las patagónicas reflejan sobrestimación (panel 1). Este comportamiento también puede verse para el modelo GFS (panel 2) lo que indica que los errores de WRF pueden responder a los errores provenientes de GFS, dado que este se utiliza como condición inicial y de borde.

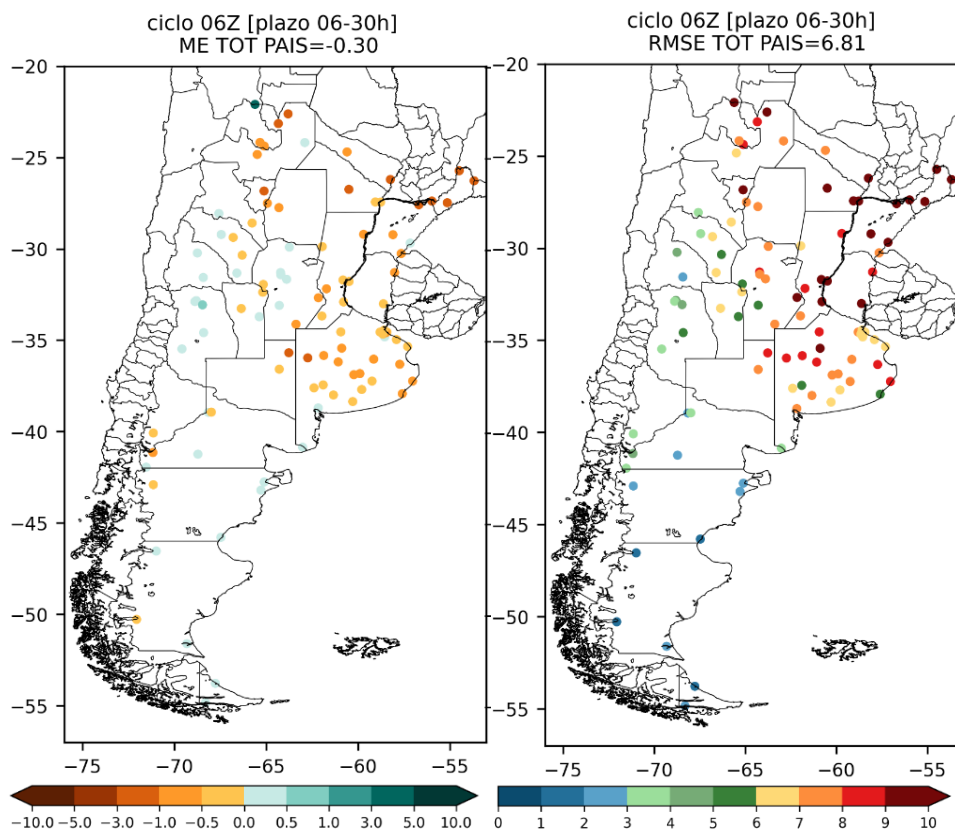


Figura 3: Mapas de BIAS de WRF (izq.) y RMSE de WRF (der.) de la precipitación diaria (ciclo 06Z, plazo 30h). Los valores en el título indican el promedio sobre el total de las estaciones sobre el periodo de verificación.

La distribución de probabilidad de volumen de precipitación (figura 4) indica que ambos modelos subestiman el volumen total tanto en los pronósticos del día 0 como los del día+1. Esto se refleja en el valor de la correspondencia entre el volumen total de precipitación observado y pronosticado por el modelo WRF es mejor que el del GFS (ver volúmenes pronosticados y medidos en el recuadro dentro de cada panel). Sin embargo, las curvas muestran que el modelo WRF acompaña mejor la forma de la distribución, siendo el modelo WRF el que explica mejor el

porcentaje de la precipitación total explicado en cada intervalo de lluvia (volumen relativo de precipitación). En cambio, la distribución del modelo GFS queda desplazada hacia las precipitaciones más débiles en comparación con el máximo de precipitación observada.

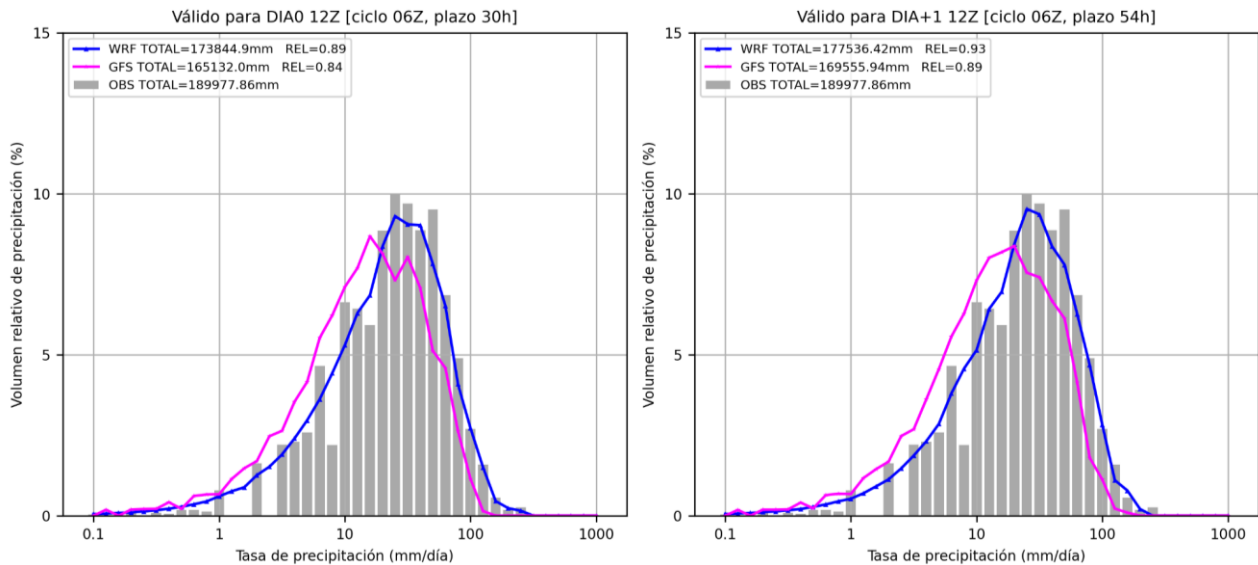


Figura 4: PDFs volumétricas de PP24 para el modelo WRF (línea azul) y GFS (línea fucsia) (ciclo 06Z) y distribución del volumen de precipitación de las observaciones (barras grises) válidas para el día 0 (izq.) y para el día+1 (der.) (ciclo 06Z). En el cuadro se indican los valores totales de precipitación observada y pronosticada considerando el total de estaciones y la relación entre el total pronosticado respecto al observado.

La figura 5 muestra que el modelo WRF tiene un mejor desempeño mostrando una relación entre la probabilidad de detección y la tasa de éxito (por sus siglas en inglés CSI, Critical Success Index) muy similar (Frequency BIAS cercanos a 1, la diagonal) en comparación con GFS para todos los umbrales. Hay que destacar que, si bien el CSI disminuye al aumentar el plazo, es poca la diferencia entre el plazo 30h y 54h.

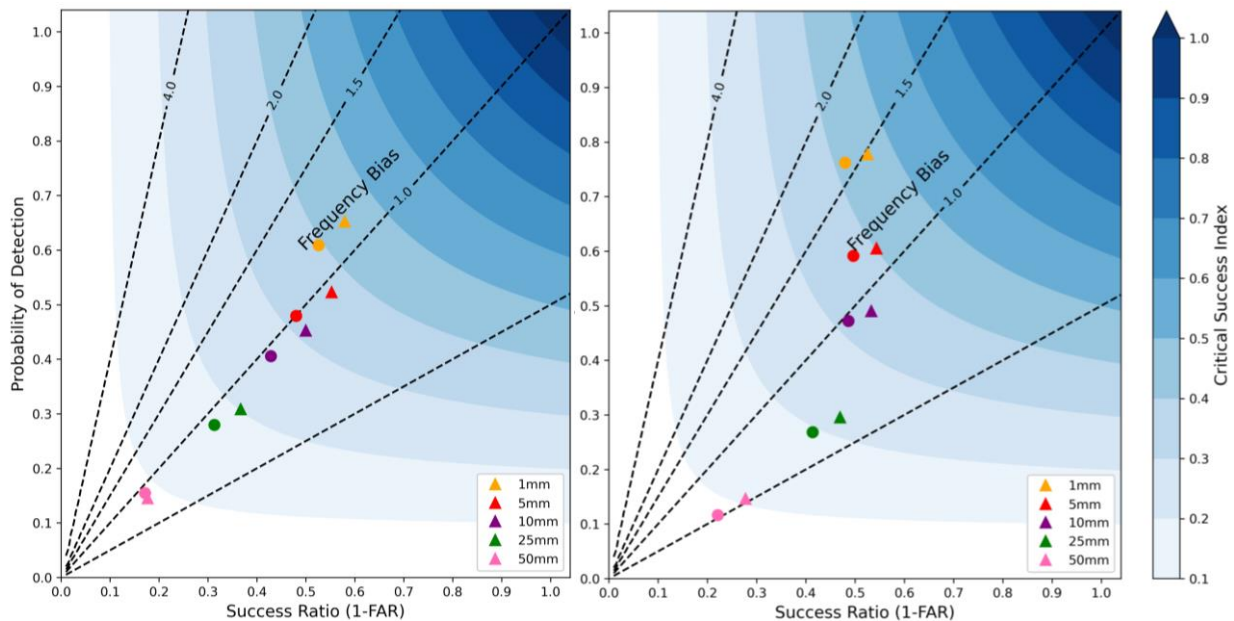


Figura 5: Diagramas de desempeño de PP24 para el modelo WRF (izq.) y GFS (der.) para distintos umbrales (ciclo 06Z) para los plazos 30h (triángulos) y 54h (círculos).

3.1.2 Temperaturas mínimas y máximas diarias

La evolución media mensual en el año de las temperaturas mínimas (T_{min_f}) y máximas diarias ($T_{máx_f}$) pronosticadas se puede ver en la figura 6. Las T_{min} pronosticadas por ambos modelos para el plazo día+1 muestran una sobrestimación para casi todos los meses del año salvo para agosto, septiembre y octubre donde se observa una leve subestimación. Los errores se encuentran entre los 2 y 3°C siendo mayores en los meses de invierno (mayo-agosto) aunque no superan los 4°C. Las T_{min} calibradas ($T_{mín_cal}$) mantienen una diferencia de error de aproximadamente 1°C respecto de los valores sin calibrar. Los valores de correlación entre los pronósticos y observaciones se mantienen superiores a 0.8 para todo el año. El porcentaje de acierto supera el 60% para las $T_{mín_cal}$ en los meses de verano.

En el caso de las temperaturas máximas, el ciclo anual se comporta en forma inversa al de las T_{min} . Esto se ve en una subestimación de los pronósticos en los meses más fríos (mar-oct) que ronda los 2°C para las temperaturas calibradas. Asimismo, es en estos meses, donde la calibración tiene más impacto lo que se refleja en los porcentajes de acierto que alcanzan valores superiores a los de la $T_{mín}$ de entre 70 y 80%.

Tanto para la $T_{mín}$ como para la $T_{máx}$, no se ven grandes diferencias para el plazo del día 2.

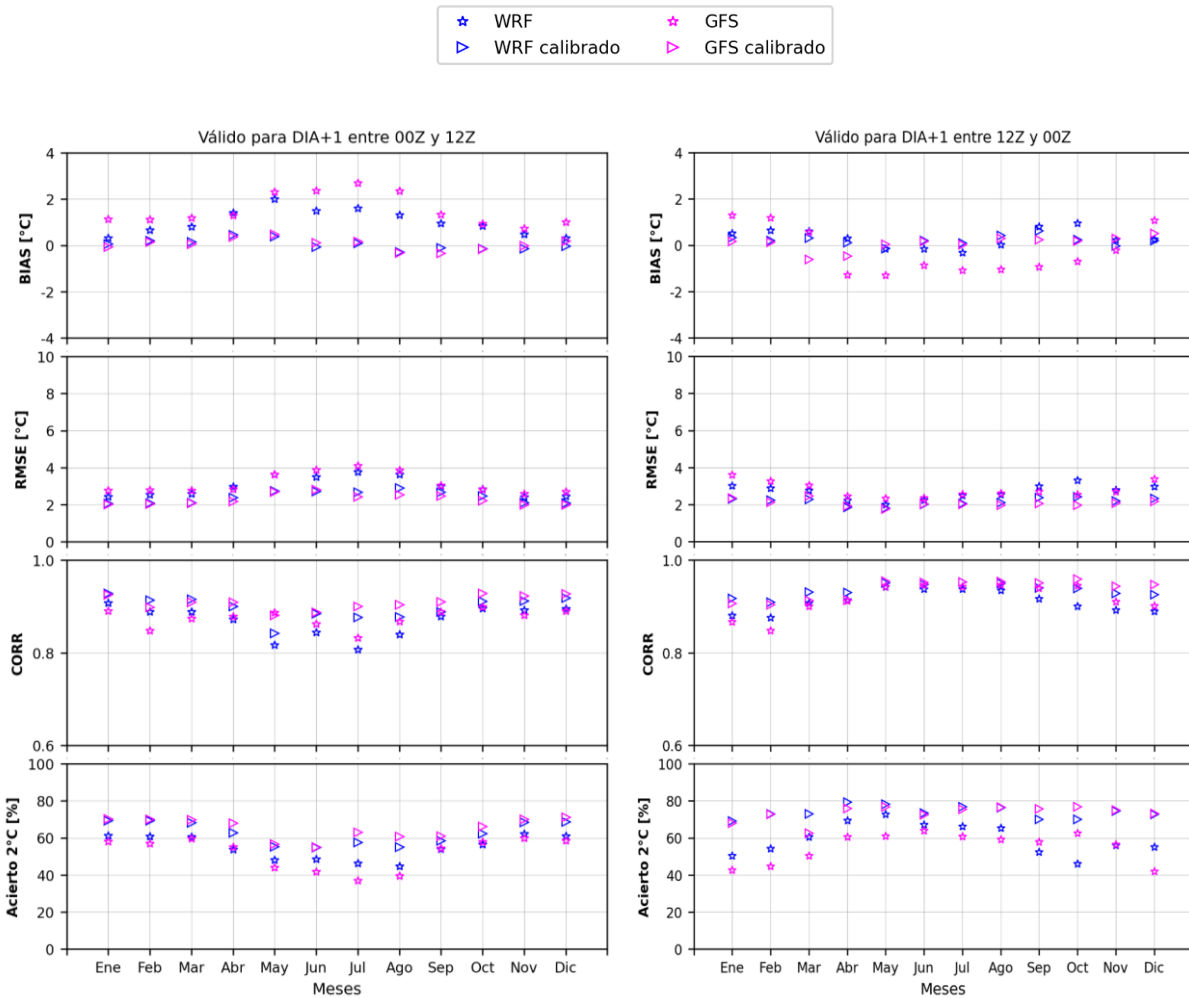


Figura 6: Evolución anual del BIAS, RMSE, el coeficiente de correlación lineal y el porcentaje de acierto para $T_{mín}$ (izq.) y $T_{máx}$ (der.) calibradas y sin calibrar (ciclo 06Z, plazo día+1).

Realizando una exploración de las diferencias entre los 4 ciclos de pronóstico, podemos ver en la figura 7 la distribución del porcentaje de acierto de todo el período para las temperaturas mínimas y máximas para el plazo día+1. Lo primero que se destaca es que el ciclo de 18Z muestra los mejores desempeños tanto para $T_{mín}$ como para $T_{máx}$ pudiendo responder a que es el ciclo que más cercano se encuentra a las observaciones. El WRF muestra mejores desempeños para ambas variables, incluso para aquellas sin calibrar, aunque se ve que la calibración tiene más impacto en las $T_{máx}$. El porcentaje de acierto de las $T_{mín_cal}$ del WRF supera en aproximadamente 6% el del GFS mientras que para $T_{máx_cal}$ esta diferencia puede llegar a alcanzar aproximadamente hasta el 12% de mejora por ejemplo para el ciclo de 12Z.

En líneas generales, para el día+2, se ve que el acierto disminuye un 4% para los ciclos de 00 y 06 y de 2% para los ciclos de 06 y 12 UTC, aunque el WRF muestra mejores desempeños que el GFS tanto para las temperaturas mínimas como máximas (no se muestra).-

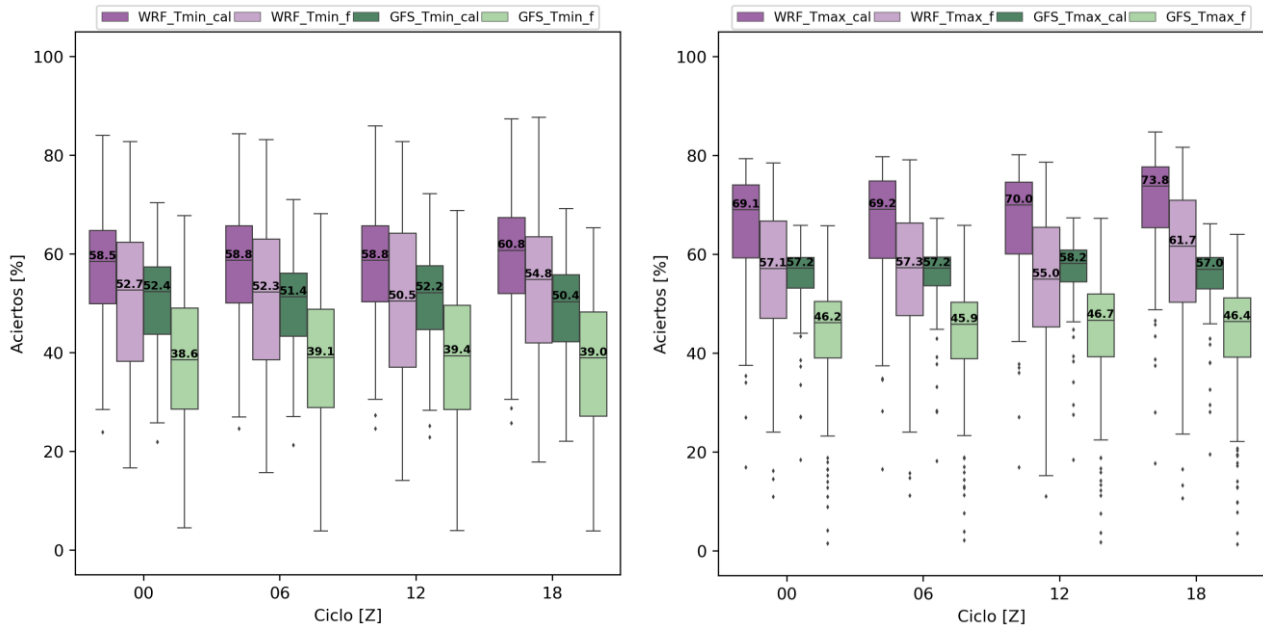


Figura 7: Boxplots del porcentaje de acierto de las Tmin (izq.) y Tmax (der.) para el plazo día+1. Los valores dentro de las cajas indican la mediana de la distribución.

La distribución espacial de estos resultados, que se puede ver en la figura 8, se ve que en el caso de las Tmin_cal, no se destaca una región con valores máximos de porcentaje de acierto, mientras que las Tmáx_cal los máximos desempeños se encuentran en la región centro del país, en la provincia de Buenos Aires y zona norte de la Patagonia. La región norte del país como las estaciones que están en zonas de topografía, los pronósticos resultan más desafiantes.

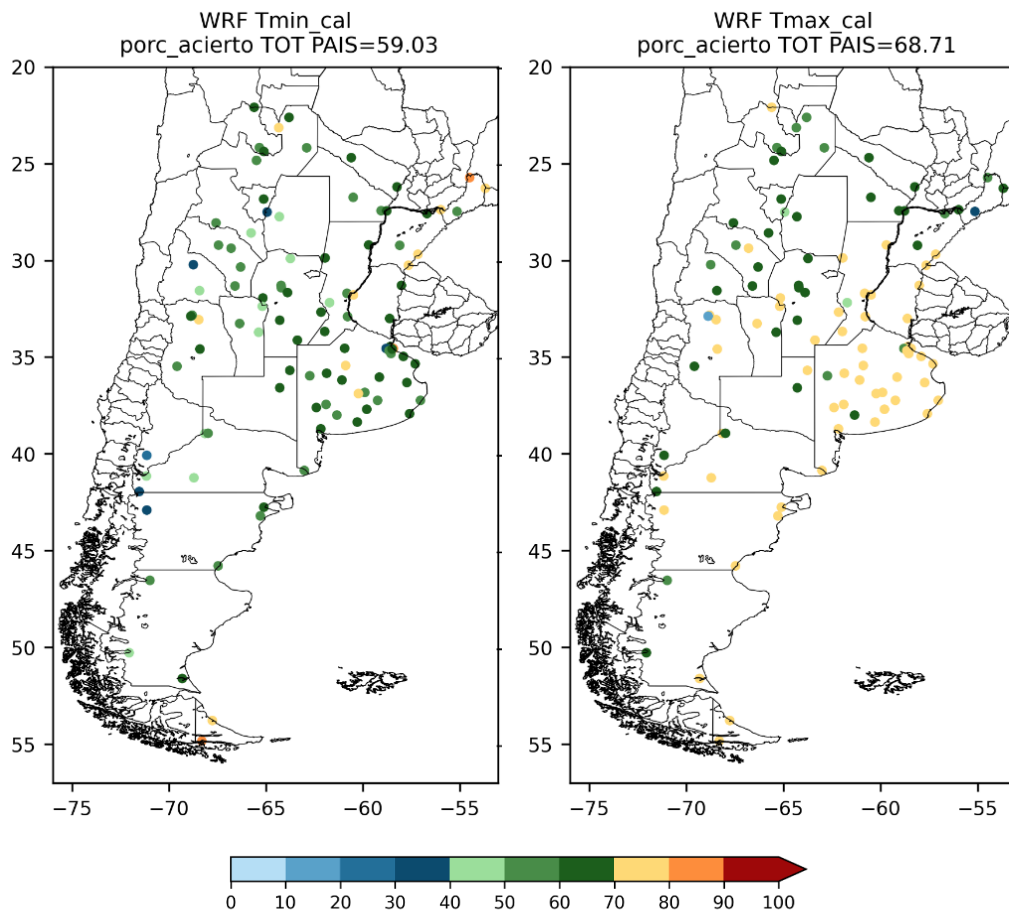


Figura 8: Mapas de porcentaje de acierto para el día+1 de la temperatura mínima calibrada (izq.) y máxima calibrada (der.) (ciclo 06Z). En el título se indica el valor del porcentaje de acierto medio para el total de estaciones.

3.1.3 Temperatura a 2m y magnitud de viento a 10m

En esta sección se muestran algunos resultados del desempeño en función del plazo de las variables horarias como temperatura a 2m y magnitud de viento a 10m calibradas. Se puede ver en la figura 9 (panel superior) que, en líneas generales, la T2cal sobrestima para todos los plazos de pronóstico con un error medio que ronda entre los 1.5 °C y 2°C hacia el final del plazo de pronóstico con BIAS máximos alrededor de las 18 UTC durante las horas de máximo calentamiento. Si bien ambos modelos se comportan de manera similar, el WRF muestra una sobreestimación levemente mayor que GFS en las horas diurnas coincidentemente con las horas donde se registran la mayor cantidad de observaciones (no se muestra). Respecto de la magnitud del viento a 10m calibrado (panel inferior), su desempeño también presenta una sobreestimación con un BIAS medio del orden de 0.1 m/s para ambos modelos y un error inferior a 1.5 m/s.

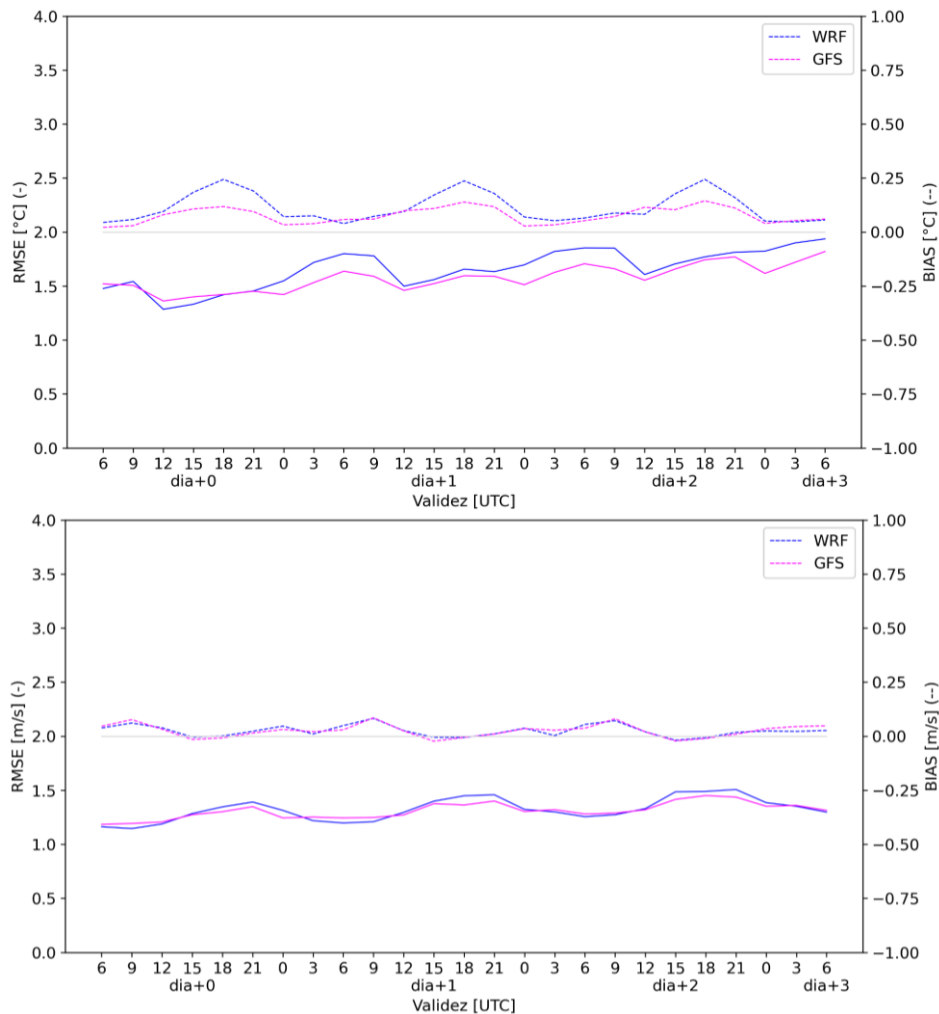


Figura 9: Serie de RMSE (eje izquierdo, líneas llenas) y BIAS (eje derecho, líneas de puntos) de la T2m calibrada (panel superior) y de magnitud del viento a 10m calibrada (panel inferior) para WRF y GFS (ciclo de 06Z).

3.2 SAP.SMN-ENS

En esta sección se analizan los resultados de la verificación de los pronósticos producidos por la componente probabilística del SAP.SMN. Se muestran diversas medidas estadísticas para el ensamble del modelo WRF y, en la medida de lo posible, se agregan resultados con el modelo GFS. Se presentan resultados que muestran el comportamiento del ensamble en su conjunto y, en algunas ocasiones, se realiza un análisis de los miembros individuales.

3.2.1 Precipitación

En relación a la verificación de los pronósticos probabilísticos de precipitación acumulada diaria (PP24), se puede ver en la figura 10, la distribución de ETS de las estaciones computadas con la media del ensamble (panel izquierdo). En sombreado se muestra el conjunto de valores de ETS calculados para todas las estaciones del país y en línea más gruesa el valor medio de esa distribución. Se puede ver que el ETS medio llega a un máximo de 0.4 para los umbrales menores a 2.5 mm para el modelo WRF. Para umbrales superiores a 25 mm también se observa un mejor desempeño del WRF frente a GFS. Por otra parte, a partir de los valores de FBIAS (panel derecho) ambos modelos sobreestiman el valor observado, aunque para umbrales menores a 25 mm, el WRF presenta un desempeño levemente mejor que el GFS.

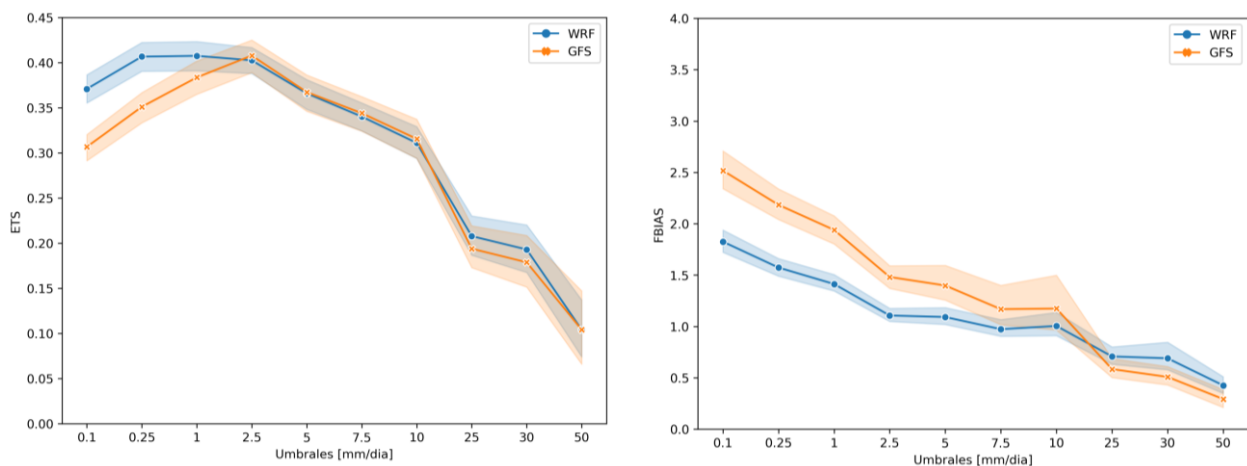


Figura 10: Serie de ETS (izq.) y Frequency BIAS (der.) de los pronósticos probabilísticos de PP24 para WRF y GFS computados con la media del ensamble para todas las estaciones se muestran en sombreado (ciclo de 06Z, válido para t+30h). La línea más gruesa indica el valor medio del conjunto de valores de los estadísticos..

A continuación, se analiza la relación entre el spread y el RMSE de la media del ensamble para los modelos WRF y GFS. En un ensamble ideal, donde todos los miembros son equiprobables, la incertidumbre en torno a la media queda representada por el spread. Por lo que se necesitan grandes (pequeñas) dispersiones para capturar eventos menos (más) predecibles donde el error RMSE entre la media y la observación es mayor. Por lo tanto, se espera que el spread y el RMSE sean comparables en cuanto a que la dispersión del ensamble logre capturar la incertidumbre de los pronósticos (Fortín, 2014). En la figura 11, la relación entre el spread y el error RMSE de la media computada para el total de estaciones muestra una mejor correspondencia para WRF que para GFS lo que se ve por el valor del coeficiente de correlación lineal de Pearson. Las estaciones con mayor spread se encuentran en el norte del país donde la precipitación media también resulta máxima (no se muestra).

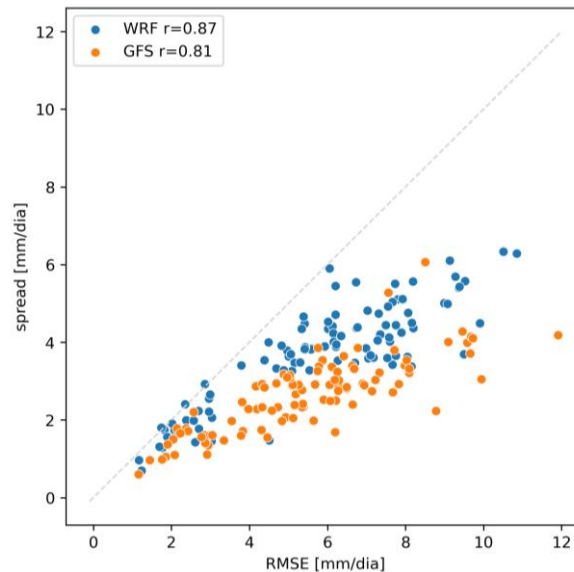


Figura 11: Diagrama de dispersión que muestra relación entre el SPREAD y RMSE de los pronósticos probabilísticos de PP24 para WRF y GFS (ciclo de 06Z, válido para t+30h). Se agrega en el recuadro el valor del coeficiente de correlación lineal de Pearson.

En la figura 12 se explora la distribución de probabilidad de volumen de precipitación de cada uno de los miembros del ensamble del SAP.SMN-ENS. Se puede ver, por los valores de relación entre el volumen total pronosticado por cada miembro respecto del total observado en los 3 años, que todos los miembros subestiman el volumen total observado. Las curvas muestran una buena correspondencia con la distribución observada. La menor dispersión ocurre para las precipitaciones más débiles mientras que para la cola derecha de la distribución, la diferencia entre los miembros es mayor.

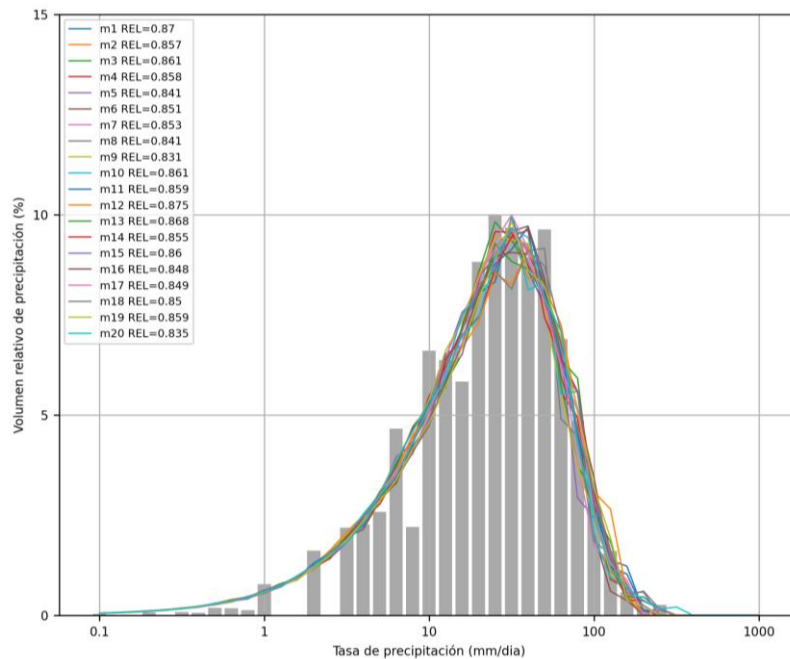


Figura 12: PDFs volumétricas de la precipitación diaria para los miembros del SAP.SMN-ENS y distribución del volumen de precipitación de las observaciones (barras grises) válidas para el día 0 (ciclo 06Z, plazo $t+30h$). En el cuadro se indican los valores totales de precipitación observada y pronosticada considerando el total de estaciones y la relación entre el total pronosticado respecto al observado para cada miembro.

El Brier score se puede considerar un índice análogo al RMSE para los pronósticos determinísticos. En ese sentido, es un indicador de la precisión del ensamble ya que se define con la magnitud media de los errores pronosticados a partir de las diferencias al cuadrado entre las probabilidades pronosticadas y las observadas. La figura 13 muestra el Brier Score en función de los umbrales. Los valores que se obtuvieron indican un buen desempeño del ensamble y son similares a los de otros sistemas de pronósticos probabilísticos operativos regionales y globales en casos de estudio de precipitación intensa de invierno en Estados Unidos (Greybush y otros, 2017).

Siguiendo a Murphy (1993, 1973), el Brier Score puede descomponerse en una componente de confiabilidad asociada al bias sistemático del ensamble (distancia a la diagonal en el diagrama de desempeño), contribuye en forma positiva al valor total de brier por lo que se busca que sea lo más pequeña posible indicando un sistema confiable. Una segunda componente es la resolución (distancia a la climatología en el diagrama de desempeño), relacionada con la habilidad del ensamble en pronosticar las probabilidades de cierto evento en comparación con la frecuencia climatológica observada. Esta componente contribuye en forma negativa al valor del brier total con valores que aumentan con el umbral debido a la disminución de las frecuencias observadas para los umbrales más grandes.

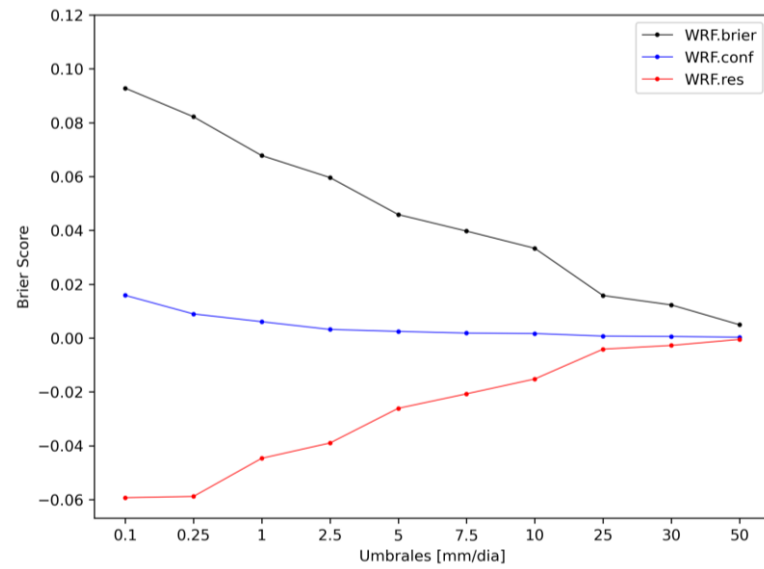


Figura 13: Brier Score total y componentes de confiabilidad y resolución de la PP24 del SAP.SMN-ENS (ciclo 06UTC, plazo $t+30h$).

El histograma de rangos resulta útil para determinar la agudeza del ensamble (Talagrand y otros, 1997; Hamill, 2001). Un ensamble perfecto mostraría un histograma “chato” donde los valores observados quedan repartidos en forma equiprobable entre los miembros. En la figura 14 (panel izquierdo) se puede ver el histograma de rangos para el SAP.SMN-ENS normalizado por la cantidad de observaciones en cada rango y donde fueron considerados los valores de precipitación superiores a 0.01 mm. Se puede observar que existe una asimetría en la distribución de las frecuencias observadas y que preferentemente, se ubican en los extremos de las clases más débiles o moderadas, mostrando que existe un bias sistemático en el conjunto y que resulta subdispersivo. Un comportamiento similar se observa con la distribución de precipitación del ensamble de GFS (no se muestra) lo que puede sugerir que las condiciones iniciales y de borde provistas por el ensamble global de menor resolución tengan un impacto directo en la representatividad de la incertidumbre del SAP.SMN-ENS.

El diagrama de confiabilidad muestra el grado de correspondencia entre las probabilidades pronosticadas de un evento con las frecuencias observadas. Es un indicador de cuán frecuente la probabilidad de un pronóstico realmente ocurrió. En el panel derecho de la figura 14 se puede ver que para los umbrales de 1 y 10 mm el comportamiento es similar donde el ensamble subestima las probabilidades bajas de hasta el 40% mientras que, para probabilidades mayores, el ensamble sobrestima las frecuencias observadas. En cambio, el umbral de 25 mm muestra una sobrestimación independientemente del valor de probabilidad. Asimismo, se ve que la curva se torna menos suave indicando que la muestra de eventos es pequeña para este umbral.

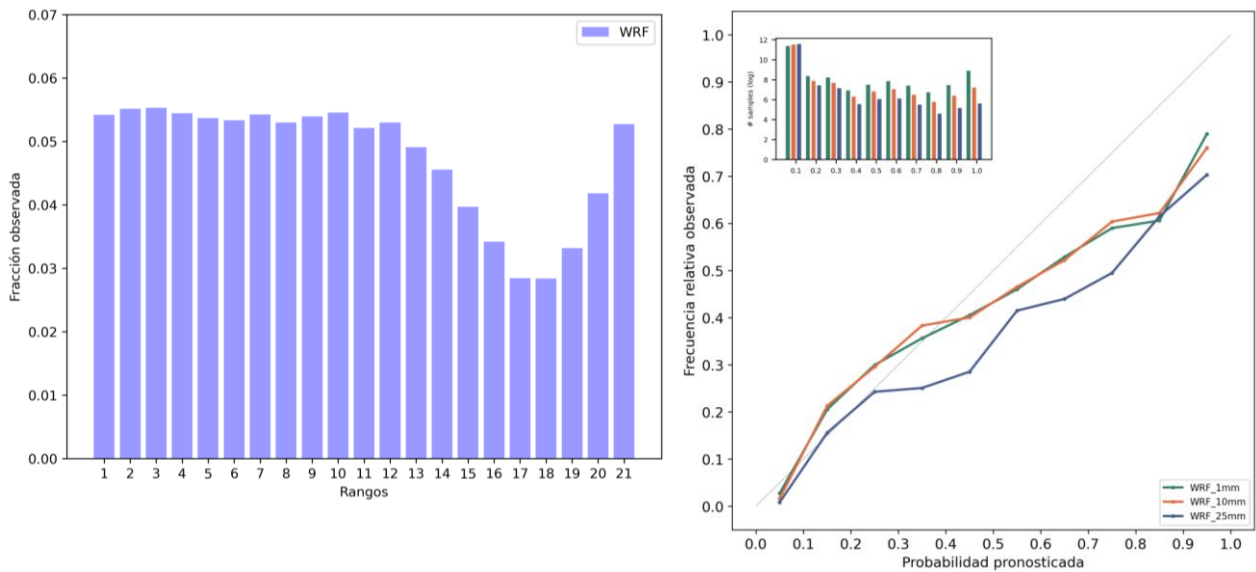


Figura 14: Histograma de rangos para la PP24 (izq.) normalizado tal que la suma de los rangos es 1. Se consideraron valores de PP24 superiores a 0.01mm y diagrama de confiabilidad (der.) para los umbrales 1, 10 y 25mm (ciclo 06Z, plazo t+30h).

3.2.2 Temperaturas mínimas y máximas diarias

Los valores medios de los aciertos de las medias del ensamble de las temperaturas mínimas y máximas pronosticadas por SAP.SMN-ENS se pueden ver en la figura 15. La calibración aumenta en aproximadamente 7% el porcentaje de acierto en el caso de las $T_{mín}$ alcanzando valores de 62.8%, mientras que en el caso de las $T_{máx}$ representa una mejora de aproximadamente 13% con valores que superan el 72%.

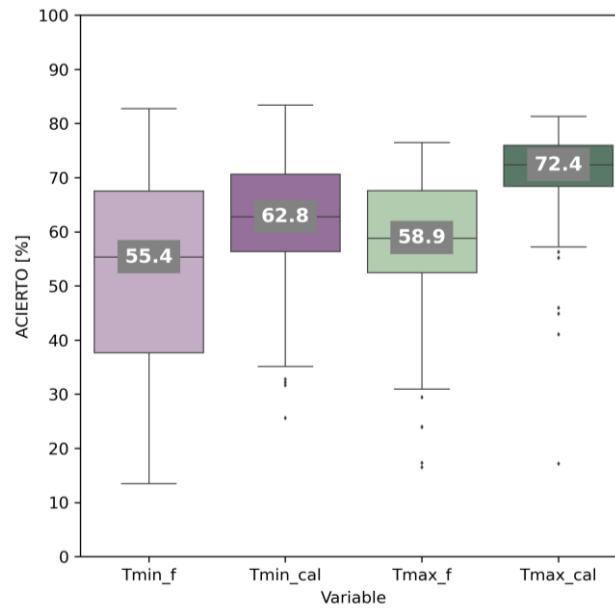


Figura 15: Boxplots del porcentaje de acierto de las Tmin y Tmáx calibradas y sin calibrar computado con las medias del ensamble. El valor indicado dentro de la caja corresponde a la mediana de la distribución (ciclo 06Z, día+1).

3.2.3 Temperatura a 2m y magnitud de viento a 10m

En cuanto al análisis de los pronósticos horarios de las variables de superficie en función del plazo podemos ver en la figura 16 la variabilidad del BIAS y la relación entre el RMSE y SPREAD. Se incluye en la figura la cantidad de observaciones disponibles en cada hora que responde a los diversos regímenes de observación en las estaciones de superficie. La T2m (panel izquierdo) muestra que, si bien la calibración reduce el RMSE y la sobreestimación, el SPREAD no alcanza magnitudes similares indicando que el ensamble sigue siendo subdispersivo. Los valores son similares para la T2 y T2cal dada la metodología de calibración en la cual la calibración de la media del ensamble se aplica a los miembros. El SPREAD toma valores máximos en horas de la tarde, lo cual podría estar asociado al comportamiento de la capa límite y, además como es de esperar, aumenta con el plazo de pronóstico. Si bien el RMSE y el BIAS muestran una variabilidad en el ciclo diario, no se ve una relación clara con el régimen de observación de las estaciones.

En el caso de la magViento10m, al igual que para la temperatura, se aprecia una sobreestimación y subdispersión por parte de los pronósticos calibrados. Se destaca la disminución del BIAS con la calibración. Asimismo, el SPREAD de la magVientocal también disminuye respecto del valor sin calibrar, posiblemente debido a la reducción en la magnitud de los valores de viento al aplicar la calibración que afecte la dispersión media.

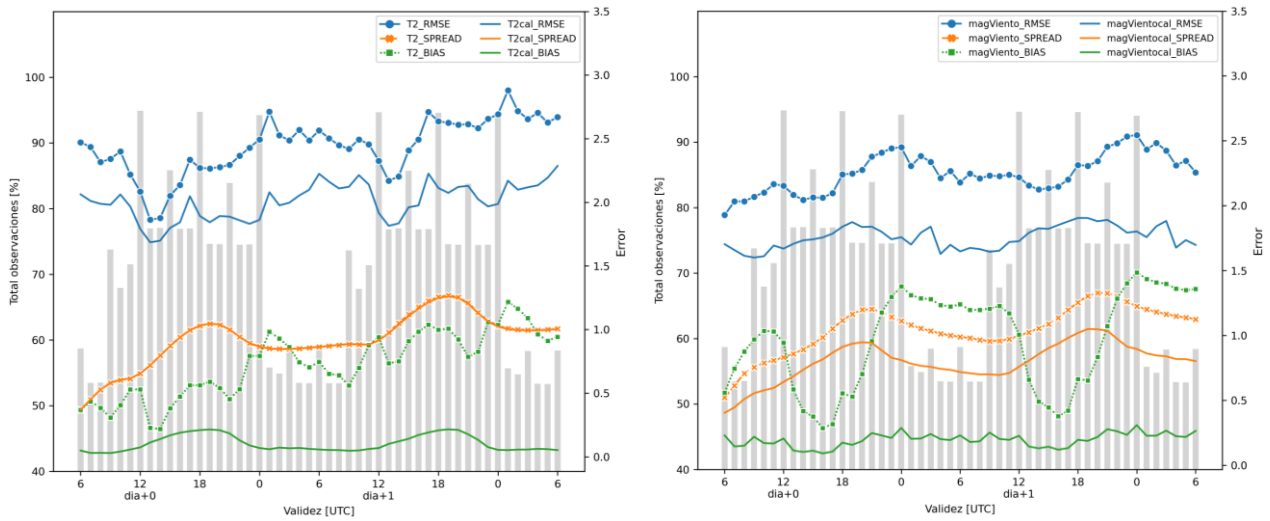


Figura 16: Series de RMSE, SPREAD y BIAS de la T2 (izq.) y de la magViento (der.) del SAP.SMN-ENS (ciclo 06Z). Las barras grises indican el porcentaje de observaciones utilizadas para la verificación para cada plazo respecto del total disponible (eje izquierdo). El eje derecho indica los valores de BIAS, RMSE y SPREAD.

4. CONCLUSIONES

Este informe recopila los resultados de la verificación anual del periodo comprendido entre los años 2020 y 2022 de los pronósticos de alta resolución generados por el sistema SAP.SMN. En particular, se analizaron las variables de precipitación acumulada en 24hs, temperatura a 2m y magnitud de viento a 10m (calibradas y sin calibrar), temperatura diaria máxima y mínima calibradas y sin calibrar. En la medida de lo posible, se realizó una comparación con el ensamble de GFS. A modo representativo se presentaron los resultados para el ciclo de 06Z ya que los otros ciclos mostraron resultados similares. Se utilizaron los datos observados provenientes de las estaciones de superficie de la red de estaciones operativas del SMN.

Respecto de la componente determinística, los pronósticos de PP24 muestran una subestimación media en el volumen total en todo el periodo, aunque el modelo WRF muestra una mejor correspondencia a la distribución de precipitación que el GFS en las tasas más altas. El frequency BIAS también muestra un mejor desempeño del WRF frente al GFS principalmente en los umbrales superiores a 5mm y el GFS sobreestima para los umbrales mayores a 25 y 50mm. En cuanto a las temperaturas mínimas y máximas diarias, el WRF muestra mejores desempeños (plazo dia+1) en los 4 ciclos, con mayor impacto de la calibración sobre la Tmax con una mejora del porcentaje de acierto del aprox 12% entre la Tmax_f y la Tmax_cal (para 12Z, plazo dia+1). Por último, la T2cal queda sobreestimada para todos los plazos de pronóstico con RMSE menores de 2°C y la magVientocal también muestra una leve sobreestimación con errores que rondan los 1.5m/s para todos los plazos.

En cuanto al desempeño del SAP.SMN-ENS, la media del ensamble de los pronósticos de PP24 muestra un mejor desempeño que el GFS para umbrales menores a 2.5 y mayores a 25 mm a pesar de que, si bien el ensamble de WRF resulta subdispersivo, se ve una mejor correspondencia entre el RMSE y el spread. Para los umbrales de 1 y 10 mm se observa una subestimación de las probabilidades mientras que para umbrales mayores a 25 mm este comportamiento se revierte. La calibración aumenta en aproximadamente 7% el porcentaje de acierto en el caso de las $T_{mín_cal}$ llegando a valores cercanos al 63%, mientras que en el caso de las $T_{máx}$ representa una mejora de aproximadamente 13% con valores superiores al 72%. Las series de RMSE y SPREAD para $T2_cal$ muestran que el ensamble resulta subdispersivo y sobreestimando los valores observados al igual que para el caso de la $magVientocal$.

Estos resultados nos alientan a continuar avanzando en el desarrollo de mejoras del SAP.SMN. En este sentido, se encuentra en proceso de implementación operativa una componente de asimilación de datos regional la cual proporcionará un avance sustancial en cuanto a la provisión de condiciones iniciales y de borde que cuenten con información actualizada de distintas fuentes locales. Asimismo, se ampliará el análisis presentado en esta nota técnica a partir del uso de otras fuentes de información observada como estimaciones satelitales de precipitación o reanálisis de otros modelos de pronóstico que permitan un estudio adecuado análisis en función de la resolución de los pronósticos.

Agradecimientos: Agradecemos al SMN y al proyecto Ciencia y Tecnología para la Producción del Alerta en Catástrofes Naturales (CyT Alerta) por la adquisición del HPC.

REFERENCIAS

Amitai, E., Petersen, W., Llord, X. y Vasiloff, S., 2012. Multiplatform comparisons of rain intensity for extreme precipitation events. *IEEE Trans. Geosci. Remote Sens.* 50 (3), 675–686. <http://dx.doi.org/10.1109/TGRS.2011.2162737>.

Cutraro F., S. Righetti, Y. García Skabar, 2022: Calibración de los pronósticos de temperatura y viento en el Servicio Meteorológico Nacional. <http://hdl.handle.net/20.500.12160/2388>

Cutraro F., S. Righetti, Y. García Skabar y M. Sacco, 2020: Implementación del sistema de pronóstico numérico en el HPC: Calibración de temperaturas pronosticadas. Nota Técnica SMN 2020-81.

Dillon, M. E., C. Matsudo, Y. García Skabar y M. Sacco, 2020a: Implementación del sistema de pronóstico numérico en el HPC: Configuración de los pronósticos determinísticos. Nota Técnica SMN 2020-78.

Dillon, M.E., Matsudo, C., Y. García Skabar, M. Sacco y M. Alvarez Imaz, 2020b: Implementación del sistema de pronóstico numérico en el HPC: Configuración del ensamble. Nota Técnica SMN 2020-79.

Fortin, V., M. Abaza, F. Ancil y R. Turcotte, 2014: Why should Ensemble Spread Match the RMSE of the Ensemble mean? *J. Hydrometeorol.* 15, 1708–1713. DOI: <https://doi.org/10.1175/JHM-D-14-0008.1>.

Greybush, S. J., Saslo, S. y Grumm, R., 2017: Assessing the ensemble predictability of precipitation forecasts for the January 2015 and 2016 East Coast winter storms. *Weather and Forecasting*, 32(3), 1057-1078. <https://doi.org/10.1175/WAF-D-16-0153.1>.

Hamill, T.M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, 129, 550-560.

Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Met. Mag.*, 30, 291-303.

Matsudo C., M.A. Salles, Y. Garcia Skabar, 2021: Verificación de los pronósticos del esquema determinístico del modelo WRF para el año 2020. Nota Técnica SMN 2021-95.

Murphy A.H, 1973: A new vector partition of the probability score. *Journal of Applied Meteorology*, 12:595–600, 1973. doi:10.1175/1520-0450.

Murphy, A.H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, 8, 281-293.

Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, 24, 601-608.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang W., Powers J., Duda M., Barker D. y X. Huang, 2019: A Description of the Advanced Research WRF Model Version 4 (No. NCAR/TN-556+STR).

Talagrand, O., Vautard R. y Strauss B., 1997: Evaluation of probabilistic prediction systems. *Proceedings, ECMWF Workshop on Predictability*.

Wilks D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, 467 p. (p 261-263).

ANEXO I: Estaciones utilizadas para la verificación

	Nombre	Lat.	Lon.	Número		Nombre	Lat.	Lon.	Número
1	BA_Aeroparque	-34.56	-58.42	87582	35	CC_Resistencia	-27.44	-59.05	87155
2	BA_Azul	-36.83	-59.89	87641	36	CC_Roque_Saenz_Pena	-26.75	-60.49	87148
3	BA_Benito_Juarez	-37.7	-59.79	87649	37	CH_Comod_Rivadavia	-45.79	-66.47	87860
4	BA_Bolivar	-36.19	-61.07	87640	38	CH_Esquel	-42.91	-71.14	87803
5	BA_Cnel_Pringles	-38	-61.33	87683	39	CH_Puerto_Madryn	-42.76	-65.1	87823
6	BA_Cnel_Suarez	-37.45	-61.89	87637	40	CR_Corrientes_Aero	-27.45	-58.76	87166
7	BA_Dolores	-36.32	-57.72	87648	41	CR_Ituzaingo	-27.59	-56.68	87173
8	BA_Ezeiza	-34.82	-58.54	87576	42	CR_Mercedes	-29.23	-58.09	87281
9	BA_Junin_Aero	-34.55	-60.93	87548	43	CR_Monte_Caseros	-30.26	-57.64	87393
10	BA_La_Plata	-34.97	-57.9	87593	44	CR_Paso_Libres	-29.69	-57.15	87289
11	BA_Las_Flores	-36.04	-59.14	87563	45	CT_Catamarca_Aero	-28.59	-65.76	87222
12	BA_Mar_del_Plata	-37.93	-57.58	87692	46	CT_Tinogasta	-28.07	-67.57	87211
13	BA_Mariano_Moreno	-34.56	-58.79	87572	47	ER_Concordia	-31.3	-58	87395
14	BA_Moron	-34.67	-58.64	87574	48	ER_Gualeguaychu	-33.01	-58.61	87497
15	BA_Nueve_Julio	-35.44	-60.88	87550	49	ER_Parana	-31.79	-60.48	87374
16	BA_Observatorio	-34.59	-58.48	87585	50	FO_Formosa	-26.21	-58.23	87162
17	BA_Olavarría	-36.89	-60.22	87643	51	FO_Las_Lomitas	-24.7	-60.59	87078
18	BA_Palomar	-34.6	-58.61	87571	52	JY_Jujuy_Aero	-24.38	-65.1	87046
19	BA_Pehuajo	-35.84	-61.86	87544	53	JY_Jujuy_Univ	-24.18	-65.33	87043
20	BA_Pigue	-37.61	-62.39	87679	54	JY_La_Quíaca	-22.1	-65.6	87007
21	BA_Punta_Indio	-35.35	-57.29	87596	55	LP_Gral_Pico	-35.69	-63.76	87532
22	BA_San_Fernando	-34.46	-58.58	87553	56	LP_Santa_Rosa	-36.59	-64.28	87623
23	BA_Tandil	-37.24	-59.23	87645	57	LR_Chemical_Aero	-30.35	-66.3	87320
24	BA_Trenque_Lauquen	-35.97	-62.73	87540	58	LR_Chepes	-31.33	-66.59	87322
25	BA_Tres_Arroyos	-38.36	-60.28	87688	59	LR_Chilecito	-29.22	-67.44	87213
26	BA_Villa_Gesell	-37.24	-57.02	87663	60	LR_La_Rioja_Aero	-29.38	-66.79	87217
27	CB_Cordoba_Aero	-31.3	-64.21	87344	61	MN_Bern_Irigoyen	-26.28	-53.67	87163
28	CB_Cordoba_Obs	-31.42	-64.2	87345	62	MN_Iguazu	-25.73	-54.48	87097
29	CB_Laboulaye	-34.13	-63.37	87534	63	MN_Obera	-27.48	-55.13	87187
30	CB_Marcos_Juarez	-32.68	-62.15	87467	64	MN_Posadas	-27.39	-55.97	87178
31	CB_Pilar	-31.67	-63.88	87349	65	MZ_Malargue	-35.48	-69.58	87506
32	CB_Rio_Cuarto	-33.1	-64.28	87453	66	MZ_Mendoza_Aero	-32.84	-68.8	87418
33	CB_Villa_Dolores	-31.95	-65.15	87328	67	MZ_Mendoza_Obs	-32.89	-68.87	87420
34	CB_Villa_Ma_RioSeco	-29.91	-63.73	87244	68	MZ_San_Martin	-33.07	-68.46	87416

	Nombre	Lat.	Lon.	Número						
69	NQ_Chapelco	-40.08	-71.14	87761						
70	NQ_Neuquen	-38.95	-68.14	87715						
71	RN_Bariloche	-41.15	-71.16	87765						
72	RN_Cipolleti	-38.94	-67.98	87719						
73	RN_EI_Bolson	-41.95	-71.53	87800						
74	RN_Maquinchao	-41.24	-68.71	87774						
75	SA_Oran	-23.15	-64.33	87016						
76	SA_Rivadavia	-24.19	-62.89	87065						
77	SA_Salta_Aero	-24.84	-65.48	87047						
78	SA_Tartagal	-22.62	-63.8	87022						
79	SC_Perito_Moreno	-46.54	-70.98	87852						
80	SC_Rio_Gallegos	-51.61	-69.31	87925						
81	SE_Rio_Hondo	-27.51	-64.93	87127						
82	SE_Sgo_del_Estero	-27.75	-64.3	87129						
83	SF_Ceres	-29.88	-61.94	87257						
84	SF_EI_Trebol	-32.19	-61.71	87470						
85	SF_Reconquista	-29.21	-59.69	87270						
86	SF_Rosario	-32.91	-60.78	87480						
87	SF_Santa_Fe	-31.71	-60.81	87371						
88	SF_Venado_Tuerto	-33.68	-61.96	87468						
89	SJ_Jachal	-30.23	-68.74	87305						
90	SJ_San_Juan	-31.57	-68.42	87311						
91	SL_San_Luis_Aero	-33.27	-66.35	87436						
92	SL_Sta_Rosa_Conlara	-32.38	-65.18	87444						
93	SL_Villa_Reynolds	-33.72	-65.37	87448						
94	TF_Rio_Grande	-53.78	-67.76	87934						
95	TF_Ushuaia	-54.84	-68.3	87938						
96	TM_Tucuman	-26.84	-65.11	87121						
97	BA_Bahia_Blanca	-38.71	-62.16	87750						
98	MZ_San_Rafael	-34.59	-68.4	87509						
99	RN_Viedma	-40.86	-63.01	87791						
100	SC_Calafate	-50.28	-72.05	87904						
101	CH_Trelew	-43.21	-65.28	87828						

Anexo II: Medidas estadísticas de verificación

Media del ensamble

$$\text{Media} = \overline{P} = \frac{\sum_{i=1}^M P_i}{M}$$

donde M es la cantidad de miembros del ensamble y P los pronósticos.

Es una medida afectada por los valores extremos. No representa a ninguno de los miembros del ensamble.

Spread medio del ensamble

$$\text{SPREAD} = \sqrt{\frac{1}{N} \sum_{k=1}^N \sum_{i=1}^M (P_{k,i} - \overline{P}_k)^2}$$

donde M es la cantidad de miembros del ensamble, \overline{P}_i representa la media del ensamble, $P_{k,i}$ es el i-ésimo miembro con $i \in \{1, \dots, M\}$ y N representa la cantidad de valores de la muestra, por ejemplo, puntos de retícula. Se denomina dispersión media del ensamble (spread) a la medida que indica la variabilidad de los miembros dada por la distancia media de los miembros al valor de la media del ensamble. Cabe mencionar que la magnitud del spread queda sujeta a los valores medios de la variable. Se utiliza en comparación con el RMSE computado con la media del ensamble para medir la incertidumbre del conjunto (Fortín, 2014).

Matriz de contingencia:

		Observado	
		NO	Total
Pronosticado	SI		
SI	Aciertos (A)	Falsas Alarmas (F)	Pronosticado si
NO	Sorpresas (S)	Correctos Neg (N)	Pronosticado no
Total	Observado si	Observado no	Total

Las categorías de la matriz de contingencia se computan en relación a valores umbrales determinados para cada variable. Por ejemplo, si se considera 1 mm como valor umbral para la precipitación, entonces si los valores pronosticado y observado son mayores a 1 mm, se considera acierto sin importar cuánta diferencia existe entre los valores de la precipitación.

Estadísticos categóricos	
<p>BIAS SCORE o Frequency BIAS (FBIAS) Indica la relación entre la frecuencia de eventos pronosticados y eventos observados. $0 \leq \text{FBIAS} \leq \infty$; valor óptimo = 1. Valores negativos indican que el modelo subestima, valores positivos indican que el modelo sobreestima.</p>	$\text{FBIAS} = \frac{A + F}{A + S}$
<p>CSI (critical success index, también llamado Threat score, TS) indica la fracción de eventos de precipitación correctamente pronosticados. No considera los eventos correctos negativos. CSI perfecto = 1, sin relación CSI = 0</p>	$\text{CSI} = \frac{A}{A + S + F}$
<p>POD (probability of detection) Indica la fracción de los eventos observados que se pronosticaron correctamente $0 \leq \text{POD} \leq 1$; valor óptimo = 1</p>	$\text{POD} = \frac{A}{A + S}$
<p>FAR (false alarm ratio) Indica la fracción de los eventos pronosticados que no se observaron $0 \leq \text{FAR} \leq 1$; valor óptimo = 0</p>	$\text{FAR} = \frac{F}{A + F}$
<p>POFD (probability of false detection) Indica la fracción de los eventos no observados que no fueron pronosticados $0 \leq \text{POFD} \leq 1$; valor óptimo = 0</p>	$\text{POFD} = \frac{F}{N + F}$

Para la verificación de la temperatura, se calcula el **porcentaje de acierto**, considerando como acierto una diferencia menor a 2 grados en el pronóstico. Por otro lado, también se calcula el **desvío o bias**, calculado como la diferencia entre el valor pronosticado y observado, es decir que un valor mayor a cero representa una sobreestimación y un valor menor que cero una subestimación.

Estadísticos continuos	
<p>En las siguientes ecuaciones, P son los pronósticos, O las observaciones, el guión representa los valores medios, y N el total de puntos considerado en el análisis.</p>	
<p>BIAS Sesgo entre el valor pronosticado y observado. Representa el error sistemático. Este índice varía entre menos infinito e infinito ($-\infty; \infty$), si el valor es mayor a 0 ($BIAS \geq 0$) indica que el modelo numérico sobrestima. Si el valor es menor a 0 ($BIAS \leq 0$) indica que subestima. El valor igual a 0 indicaría que la observación es igual al valor pronosticado para esa variable.</p>	$BIAS = \frac{\sum_{i=1}^N (P_i - O_i)}{N}$
<p>RMSE (root mean square error) Medida de las diferencias cuadráticas del valor medio de los errores de los pronósticos. En el caso del ensamble, se computa utilizando la media del ensamble o los miembros en forma individual. Su valor varía entre cero e infinito ($0; \infty$).</p>	$RMSE = \sqrt{\frac{\sum_{i=1}^N (P_i - O_i)^2}{N}}$
<p>Coefficiente de correlación lineal de Pearson Representa el grado de asociación lineal entre pronósticos y observaciones. Su valor varía entre cero y uno $[0;1]$, donde 0 indica que no existe relación lineal entre las variables y 1 que la relación entre ellas es perfectamente lineal.</p>	$r = \frac{\sum_{i=1}^N (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}$

Distribución de probabilidad volumétrica de precipitación (PDF)

Las distribuciones de probabilidad del volumen de precipitación (Amitai y otros, 2012), subdividen la tasa de precipitación diaria en intervalos en escala dBR, y consideran la contribución relativa de los mismos al volumen total de precipitación. Se definen como la suma de las tasas de precipitación para un dado intervalo en escala dBR ($dBR = 10 \log(R/1 \text{ mm día}^{-1})$; $[R] = [\text{mm día}^{-1}]$) dividido por la suma total de las tasas de precipitación. La ventaja respecto a las PDFs de ocurrencia, es que son menos sensibles a las limitaciones en la detección de precipitación débil (asociada a una pequeña fracción de la precipitación total). En la leyenda de las figuras, se incluye la relación entre la precipitación total pronosticada y observada sobre la región y período analizado. Estas figuras se muestran para la PP24 pronosticada con el ciclo 06Z y válida para el día 0 y día+1 a las 12Z.

Diagrama de desempeño

El diagrama de desempeño constituye una forma geométrica de mostrar la relación entre las medidas de verificación categóricas: la probabilidad de detección (Probability of Detection, eje y en el diagrama), la tasa de falsas alarmas (FAR, de sus siglas en inglés) o su opuesto la tasa de éxito (Success Ratio = 1-FAR, eje x en el diagrama), el Frecuency Bias (líneas punteadas) y el índice crítico de éxito (Critical Success index, valores sombreados en el diagrama). Más detalles sobre la construcción de este diagrama pueden encontrarse en Roebber (2009).

Un buen desempeño de los pronósticos se refleja en valores de POD, SR, FBIAS y CSI cercanos a 1, por lo que un pronóstico perfecto se encontrará en la esquina superior derecha del diagrama. Las desviaciones en alguna dirección indican diferencias relativas entre POD y SR y consecuentemente entre FBIAS y CSI. La ventaja de este diagrama es que las diferencias en el desempeño quedan visibles en forma inmediata.

Histograma de rangos (o diagrama de Talagrand)

El histograma de rangos es una representación de la confiabilidad del sistema de pronóstico por ensamble en el cual las categorías asociadas a cada barra son rangos variables en lugar de valores discretos. La forma del diagrama nos permite sacar conclusiones sobre los sesgos sistemáticos y el alcance de su dispersión. Un histograma plano en el cual todas las columnas son iguales indica una confiabilidad perfecta. Otros patrones brindan información sobre los tipos de sesgos y las dispersiones imperfectas (vea el apartado sobre interpretación, más adelante).

El eje x representa las clases cuyos valores mínimos y máximos quedan definidos por los miembros mismos del conjunto, clasificados del valor más bajo al más alto (en otras palabras, las clases son los rangos entre los miembros de pronósticos por conjuntos). Cada celda de la malla en cada dominio del pronóstico tiene un conjunto exclusivo de clases definidas de esta manera para cada pronóstico. El eje horizontal no contiene el valor de los rangos, sino que indica simplemente que se utilizaron los mismos rangos de límites de clase en cada pronóstico. Por lo que el valor de los rangos puede variar bastante. Si dos miembros del conjunto tienen el mismo valor, sus clases se combinan. Por ejemplo, si dos miembros del conjunto tienen el mismo valor (por decir, 5650 metros) en determinado punto del ciclo de ejecución del conjunto y fueron clasificados como 9 y 10, después de la clase 8, la clase 9 se combinaría con la clase 10. No habría ninguna contribución a la clase 9 para ese punto y en ese momento. La primera clase representa el rango hasta el miembro del conjunto de la clase más baja, mientras que la última clase representa el rango por encima del miembro del conjunto en la clase más alta. Esto significa que hay $n+1$ clases en el diagrama siendo n la cantidad de miembros.

Las frecuencias mostradas por las barras representan el porcentaje de veces que las observaciones caen en los rangos o las clases definidas por los miembros de la predicción por conjuntos en cada punto de retícula del dominio dentro del período de verificación. La primera clase contiene las verificaciones que son menores que el miembro del conjunto en la clase más baja, mientras que la última clase contiene las verificaciones que son iguales o más altas que el miembro del conjunto en

la clase más alta. Si una observación es igual a uno de los miembros del conjunto, esta se divide entre las dos clases adyacentes.

Índice de Brier (BS) y componentes

El Brier Score es una medida de los errores de los pronósticos probabilísticos. Se basa en el promedio de las diferencias al cuadrado de las probabilidades pronosticadas y las frecuencias observadas de los pares pronóstico-observación:

$$BS = \frac{\sum_{i=1}^N (f_i - o_i)^2}{N}$$

donde f son las probabilidades pronosticadas y o es la frecuencia observada, N es el número total de casos. El BS toma valores de 0 (pronósticos perfectos) a 1 (perfectamente erróneos). Es una medida que indica la precisión de los pronósticos (Murphy, 1973).

Se puede descomponer en componentes (véase, por ejemplo, Wilks 1995, 261-263) como:

BS = confiabilidad - resolución + incertidumbre

$$BS = \frac{1 \sum_{i=1}^B n_i (f_i - o_i)^2}{N} - \frac{1 \sum_{i=1}^B (o_i - c)^2}{N} + c(1 - c)$$

donde f_i son los pronósticos probabilísticos en el intervalo de probabilidad i , o_i es la frecuencia observada en el intervalo de probabilidad i , N es el número total de casos, B es el número de intervalos de probabilidad, n_i es el total de casos en el intervalo de probabilidad i , c es la frecuencia observada en toda la muestra.

La confiabilidad depende del cuadrado de las diferencias entre las probabilidades pronosticadas y las frecuencias observadas condicionales. Mide el sesgo entre la probabilidad pronosticada y la frecuencia media observada para un umbral e intervalo de probabilidad (distancia a la diagonal en el diagrama de confiabilidad), mide bias sistemático del ensamble, es la componente que se podría calibrar. Se busca que el valor sea lo más bajo posible.

La resolución se calcula con las probabilidades pronosticadas condicionadas a la frecuencia observada para ese umbral e intervalo de probabilidad (distancia a la línea horizontal de la climatología en el diagrama de confiabilidad). Se busca que sea lo más grande posible de modo de que las frecuencias observadas sean lo más distintas a la climatología.

El término de incertidumbre mide la varianza de la muestra de observaciones por lo que es independiente del sistema de pronóstico por lo que resulta independiente de cualquier calibración aplicada. Es el eje y del diagrama de confiabilidad (es la frecuencia observada para cierto umbral). Toma valores entre 0 y 0.25. Es 0 si el evento ocurre siempre o nunca, y se acerca a 0.25 cuando la probabilidad climatológica es cercana a 0.5 hay más incertidumbre inherente al pronóstico. Cuanto más cercana la probabilidad climatológica de los valores extremos (ocurre siempre o no ocurre nunca), más baja será la variabilidad y la contribución al BS se reduce al mínimo. Si la probabilidad de que se produzca un evento es igual a la probabilidad de que no se produzca (probabilidad

climatológica de 0.5), la variabilidad de las observaciones y la contribución de la incertidumbre al índice de Brier se maximiza.

Diagrama de confiabilidad (“diagrama de atributos”)

El diagrama de confiabilidad muestra las probabilidades pronosticadas en función de las frecuencias observadas donde el rango de los valores de los pronósticos se divide en B intervalos de probabilidad (por ejemplo, 0-5%, 5-15%, 15-25%, etc.). La cantidad total en cada intervalo se muestra como un gráfico de barras. Mide la correspondencia entre las probabilidades pronosticadas de un evento y las frecuencias observadas.

La confiabilidad queda representada por la proximidad de la curva a la diagonal. La distancia a la curva indica el bias condicional. Si la curva se encuentra por debajo de la diagonal indica que el pronóstico sobrestima (las probabilidades son muy altas); mientras que si la curva queda por encima indica subestimación (las probabilidades son muy bajas). Cuanto más chata es la curva, menos resolución tiene. Un pronóstico climatológico no discrimina la ocurrencia o no de eventos por lo que su resolución es cero. La frecuencia de los pronósticos en cada intervalo de probabilidad (en el histograma) muestra la agudeza del sistema de pronóstico. Cabe destacar que este diagrama queda condicionado a la muestra de pronósticos por lo que se recomienda acompañarlo con el diagrama ROC el cual está condicionado por las observaciones.

Diagrama Relative Operating Characteristics (ROC)

El diagrama ROC (Mason, 1982) muestra la tasa de aciertos (POD) en función de la tasa de falsas alarmas (POFD) usando un conjunto de umbrales de probabilidad ascendente (por ejemplo, 0.05, 0.15, 0.25, etc.) para poder computar la tabla de contingencia. Sirve para responder cuál es la habilidad del pronóstico para discriminar entre la ocurrencia o no de eventos (resolución). Una curva de ROC perfecta va desde la esquina inferior izquierda hacia la esquina superior izquierda y luego hacia la esquina superior derecha. La diagonal indica no habilidad. El área bajo la curva de ROC toma valores desde 0 a 1 (pronóstico perfecto), 0.5 indica que no hay habilidad. No es sensible al bias por lo que no indica nada sobre la confiabilidad. Un pronóstico con bias sistemático puede tener buena resolución y mostrar una buena curva de ROC por lo que podría mejorar mediante una calibración. El ROC está condicionado por las observaciones (dado un evento que ocurrió, ¿cuál fue el pronóstico correspondiente?) por lo que debe analizarse junto con el diagrama de confiabilidad. Se puede considerar como una medida de utilidad potencial de los pronósticos.

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 Martin 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).