

Verificación de los pronósticos del esquema probabilístico del modelo WRF para el año 2020

Nota Técnica SMN 2022-120

Cynthia Matsudo¹, María Alejandra Salles¹ y Yanina García Skabar^{1,2}

¹Dirección de Productos de Modelación Ambiental y Sensores Remotos, Dirección Nacional de Ciencia e Innovación en Productos y Servicios, Servicio Meteorológico Nacional ²Consejo Nacional de Investigaciones Científicas y Técnicas de Argentina (CONICET)

Enero 2022





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

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 operativos del modelo WRF del esquema probabilístico correspondientes al año 2020. Las variables que se verifican son: temperatura a 2m, 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. Solo para el caso de las temperaturas máximas y mínimas diarias se comparan los resultados con los correspondientes al ensamble del Global Forecast System (GEFS). En líneas generales, el ensamble resulta subdispersivo, el análisis del desempeño de los 4 ciclos de pronóstico arrojó resultados similares. En cuanto a la precipitación acumulada en 24 hs, la media del ensamble mostró una subestimación similar que el esquema determinístico aunque una inspección sobre los miembros reveló que tienen una mejor correspondencia con la distribución de frecuencias observada. El Brier Score resulta similar 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. En cuanto a las temperaturas horarias, la calibración disminuye los errores aproximadamente en un 18%. El impacto de la calibración se ve más claramente en el caso de las temperaturas mínimas diarias. El desempeño del ensamble WRF es mejor que el del GEFS tanto para los valores calibrados como sin calibrar. Se evidencian diferencias en la calidad del pronóstico de algunas variables al utilizar determinadas combinaciones de parametrizaciones. Los resultados obtenidos pueden responder a la cantidad de miembros del sistema, a la elección de la configuración física, así como la metodología empleada en el cálculo de los estadísticos relacionada con la agregación de los valores pronosticados como considerar distintas climatologías, entre otros factores.

Abstract

This technical note is developed within the framework of the National Meteorological Service (NMS) Forecast Cross-Verification Plan. Here are presented the operational forecast verification results of the WRF model of the probabilistic scheme corresponding to the year 2020. The variables that are verified are the following: 2-m temperature, daily minimum and maximum temperatures and 24hour accumulated precipitation. Observations for verification come from the NMS surface stations network. Only for daily minimum and maximum temperatures a comparison with Global Ensemble Forecast System (GEFS) forecasts was made. Overall, the ensemble is underdispersive, the 4-cycle comparative analysis revealed similar results. In terms of 24-h accumulated rainfall, the ensemble mean was underestimated similarly to what was found for the deterministic scheme although an inspection on the ensemble members revealed that they have a better correspondence with the observed frequency distribution. Brier Score is similar to other regional ensemble forecast systems,





and from reliability values a systematic bias is detected for which a calibration process could be required. Calibrated hourly temperatures show a skill improvement of approximately 18%. The calibration impact is more clearly seen for daily minimum temperatures. The WRF ensemble performance is better than for the GEFS for calibrated and uncalibrated values. Looking at the ensemble members there were some differences in skill for some parametrization combinations. Results can respond to the number of ensemble members, the selection of physics configuration, to the applied verification metrics and the aggregation of different climatologies among other factors.

Palabras clave: Verificación, ensamble, WRF, GFS

Citar como: Matsudo C., M.A. Salles, Y. Garcia Skabar, 2022: Verificación de los pronósticos del esquema probabilístico del modelo WRF para el año 2020. Nota Técnica SMN 2022-120.





1. INTRODUCCIÓN

Un pronóstico por ensambles es un conjunto de pronósticos que presenta un rango de futuros estados de la atmósfera. Se realizan múltiples simulaciones, cada una a partir de una pequeña variación en las condiciones iniciales y/o en la configuración física de los modelos las cuales representan la incertidumbre inevitable de las condiciones iniciales y la aproximación de los modelos. La incertidumbre asociada a cada pronóstico indica que son posibles distintos escenarios a diferencia del pronóstico determinístico que no provee esa información. Con la generación de un rango de escenarios posibles, el método logra mostrar el nivel de probabilidad con el que se cumplan y hasta qué plazo en el futuro esos pronósticos resultan útiles. Cuánto más pequeño es el rango de los posibles resultados, el pronóstico logra ser más 'preciso'. Los ensambles de buena calidad no solo son precisos sino también confiables. La incertidumbre está asociada a la falta de conocimiento de las condiciones iniciales y a la aproximada representación de los procesos atmosféricos en los modelos numéricos además de a la naturaleza caótica intrínseca de la atmósfera. 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 los pronósticos por ensambles la incertidumbre está dada por la dispersión del conjunto.

En enero de 2020, el SMN comenzó a producir pronósticos probabilísticos de alta resolución con el modelo WRF junto con el esquema determinístico en el sistema de cómputo de alta performance HPC, Huayra Muyu (HM). Se realizaron múltiples pruebas de configuración del ensamble las cuales implicaron la evaluación de cantidad de miembros, configuraciones físicas, tiempos de procesamiento, versiones del modelo, entre otras cosas. Los detalles de la configuración final del ensamble multifísico de 20 miembros que se ejecuta en forma operativa se pueden encontrar en Dillon y otros (2020).

Todo sistema de pronóstico numérico debe contar con un esquema de verificación para comprobar y mejorar 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 Matsudo y otros (2021). Para las predicciones por conjuntos, también se pueden aplicar estas y otras herramientas estadísticas típicas al valor medio de una variable meteorológica o al valor de un miembro individual. Sin embargo, para los pronósticos probabilísticos, no deseamos evaluar solo la exactitud de un miembro individual o de la media del conjunto sino también verificar la distribución de probabilidad de los eventos posibles en cuanto a cómo las probabilidades pronosticadas se asemejan a la





frecuencia de los eventos observados o la utilidad de los pronósticos en relación a la climatología y a los pronósticos de referencia como puede ser la corrida determinística.

El objetivo de esta Nota Técnica es presentar una síntesis del desempeño de los pronósticos probabilísticos para el año 2020 de precipitación acumulada en 24hs y temperatura a 2m, temperatura mínima y máxima sin calibrar y calibrada generados por el modelo WRF. En el caso del análisis de las temperaturas diarias, se realiza una comparación con los pronósticos del ensamble de GFS de menor resolución.

2. DATOS Y METODOLOGÍA

Para este trabajo se utilizan los pronósticos horarios del modelo WRF-ARW (versión 4.0) que se generan en forma operativa en los ciclos 00, 06, 12 y 18 UTC en el SMN y forman parte del esquema probabilístico que consta de un ensamble de 20 miembros que combinan distintas configuraciones de microfísica y capa límite (ver Tabla I). Estos pronósticos tienen una resolución horizontal de 4 km, 45 niveles verticales (tope 10 hPa) y plazo máximo de 48 hs. Utilizan como condición inicial y de borde, los análisis y pronósticos del ensamble de GFS como se describe en Dillon y otros (2020). Las dos configuraciones de microfísica de un momento que fueron utilizadas son WRF Single Moment 6 (WSM6) y Lin mientras que para capa límite se utilizaron Mellor-Yamada-Janjic (MYJ), Yonsei University (YSU) y Shing Hong (SH).

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	

Tabla I. Combinación de las configuraciones físicas en el ensamble.

Las variables pronosticadas que se verifican son: temperatura a 2 metros (T2m), temperatura a 2 metros calibrada (T2cal), temperatura mínima y máxima calibrada y sin calibrar (Tmín, Tmáx, Tmíncal, Tmáxcal respectivamente) y precipitación acumulada en 24 horas (PP24).





Para todos los ciclos, la Tmín corresponde al valor mínimo horario pronosticado entre 00 y 12Z (horas de la mañana) mientras que la Tmáx 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). Los datos observados de precipitación corresponden a la precipitación acumulada en 24 horas válida a las 12Z de cada día (día pluviométrico).

Los datos observados provienen de la red de estaciones de superficie del SMN disponible en tiempo real. Por cuestiones metodológicas y de disponibilidad de información, para este informe se consideraron 101 estaciones meteorológicas (ver ANEXO I).

Para el cálculo de los estadísticos (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 considera 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 las distintas variables.

3.1 Precipitación

En primer lugar se hace una inspección del desempeño de la media del ensamble abordando una evaluación similar a la que se realizó para el esquema determinístico. Dentro de las ventajas que presenta el uso de la media del ensamble se encuentra el concepto de que filtra las escalas más pequeñas con poca predictibilidad y la interpretación de los resultados resulta más fácil para los usuarios que los estadísticos probabilísticos. Sin embargo, cabe destacar que no forma parte de ninguna de las simulaciones del ensamble por lo que hay que considerar que las propiedades estadísticas son diferentes para el ensamble y para las observaciones.

En todos los ciclos de ejecución de los modelos (00Z, 06Z, 12Z y 18Z) se observan características similares, a modo de ejemplo se muestra el ciclo de las 06Z para el cual la verificación se aplica a la hora de plazo 30h.

En primer lugar, se puede ver en la Figura 1 el ETS y BIAS Score de la media del ensamble. Estos muestran resultados similares al desempeño del esquema determinístico para el ciclo de 06Z (no mostrado). En el caso del BIAS Score se puede ver un mejor desempeño para los umbrales menores a 5 mm que el esquema determinístico.







Figura 1: ETS (*izq.*) y BIAS Score (der.) computados con la media del ensamble en función de distintos umbrales de precipitación acumulada en 24hs para el ciclo 06Z.

En el caso de la distribución de probabilidad de volumen de precipitación (Figura 2) se puede ver, en los valores en el recuadro, la correspondencia entre el volumen total anual de precipitación observado y pronosticado por cada uno de los miembros y por la media del ensamble. Por otra parte, las curvas indican el porcentaje de la precipitación total explicado en cada intervalo de lluvia (volumen relativo de precipitación). Si bien todos los miembros y la media subestiman el volumen total, se puede ver que los miembros muestran una mejor correspondencia con las observaciones que la media la cual muestra una distribución desplazada hacia las precipitaciones más débiles en comparación al máximo observado. Lo cuál es esperable ya que la media reduce los valores extremos.







Figura 2: PDF volumétrica de la media (línea negra) y los miembros del ensamble de la precipitación diaria para el modelo WRF (ciclo 06Z) y distribución del volumen de precipitación de las observaciones (barras grises). En el cuadro se indica para cada miembro y la media del ensamble la relación entre el total pronosticado respecto al observado. En la Figura 3 de Matsudo y otros (2021) puede verse el mismo gráfico para la corrida determinística.

Por último, se muestra el diagrama de desempeño (Figura 3) que permite ver los valores de distintos estadísticos categóricos como el BIAS Score, el POD, FAR y CSI computados con la media del ensamble para distintos umbrales. Se pueden ver valores más altos de CSI en comparación con el esquema determinístico lo que indica una mayor cantidad de eventos correctamente pronosticados para todos los umbrales. Sin embargo, al igual que en esquema determinístico, se subestima la cantidad de eventos con umbral menor a 1 mm y se mantiene la sobrestimación para los otros umbrales.







Figura 3: Diagrama de desempeño de la precipitación diaria computado con la media del ensamble para el modelo WRF para distintos umbrales (ciclo 06Z) para la hora de plazo 30h. En la Figura 4 de Matsudo y otros (2021) puede verse el mismo gráfico para la corrida determinística.

De manera exploratoria, para el ciclo de 06Z, se realizaron los diagramas de desempeño utilizando la media del ensamble para los distintos subconjuntos de parametrizaciones con el objetivo de analizar las diferencias entre ellos. Los valores de los estadísticos resultaron similares en todos los umbrales (no se muestra) por lo que no se pudo detectar una sensibilidad a las parametrizaciones físicas.

Una de las maneras para caracterizar el ensamble es mediante el uso del histograma de rangos (Talagrand et al., 1997; Hamill, 2001) donde se mide la frecuencia de las observaciones en función de rangos de valores pronosticados por el ensamble. Es una herramienta de diagnóstico con la cual se puede medir la calidad del ensamble respecto de si tiene la suficiente dispersión como para capturar la variabilidad observada. En este caso, se muestra el histograma para el ciclo de 06Z normalizado por la cantidad de observaciones en cada rango y donde fueron considerados los valores de precipitación superiores a 0.01mm. Se puede ver que existe una asimetría en el histograma lo que indica un bias en el ensamble que concuerda con los resultados obtenidos por las PDFs (Figura 2). Se ve que la frecuencia de las observaciones es más alta de lo esperado en las clases bajas y moderadas y más bajas de lo esperado en los rangos más altos de los pronósticos. Esto sugiere que los valores pronosticados por los miembros se ubicaron en el centro de la muestra otorgando un grado de dispersión más bajo que el necesario para capturar la variabilidad de las





observaciones. Esto puede deberse a la variación en la configuración física, al tamaño del ensamble y al tamaño de la muestra observada. Los demás ciclos de pronóstico muestran resultados similares.



Figura 4: Histograma de rangos para la PP24 (ciclo 06Z) normalizado tal que la suma de los rangos es 1. Se consideraron valores de PP24 superiores a 0.01mm.

A continuación se presentan los resultados que muestran distintas cualidades del desempeño del ensamble. El Brier Score indica la precisión del ensamble a partir de computar las diferencias al cuadrado entre las probabilidades pronosticadas y las observaciones, por lo que otorga mayor peso a los errores más grandes, cuánto más chico resulta el valor, mejor el desempeño del ensamble. Es el estadístico análogo al RMSE para los pronósticos determinísticos por lo que da idea de la magnitud de los errores de los pronósticos probabilísticos. 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 et al., 2017).

El Brier Score se puede descomponer en 3 componentes que ayudan a interpretar las fuentes de los errores en los pronósticos (Murphy 1973) lo que resulta útil a la hora de explorar la dependencia entre las probabilidades pronosticadas y las características del ensamble.

La confiabilidad es una medida del sesgo entre la probabilidad pronosticada y la frecuencia media observada para cada umbral e intervalo de probabilidad (o sea, la distancia a la diagonal en el diagrama de confiabilidad) por lo que mide el bias sistemático del ensamble. Buscamos que su valor





sea lo más pequeño posible de modo de que el ensamble resulte confiable lo cual sucede en nuestro caso donde contribuye en una pequeña proporción al valor del brier total.

La resolución mide la diferencia de las frecuencias observadas condicionales y la frecuencia climatológica para intervalo de probabilidad pronosticada por lo que se busca que su valor sea grande de modo que las frecuencias observadas de la muestra queden apartadas de la climatología. Esto es lo que sucede en nuestro caso donde es alta la contribución de esta componente para los umbrales bajos mientras que para a medida que aumenta el umbral se observa que queda condicionada a la baja disponibilidad de días con precipitación en las estaciones de observación.



Figura 5: Brier Score (línea negra) y componentes de confiabilidad (línea azul), resolución (línea roja) de la precipitación diaria para el modelo WRF (ciclo 06Z).

En la Figura 6 el diagrama de confiabilidad muestra el grado de correspondencia entre las probabilidades pronosticadas de un evento con las frecuencias observadas. Mide cuán frecuente la probabilidad de un pronóstico realmente ocurrió. Tiene mayor dependencia de los pronósticos que de las observaciones. A modo de ejemplo, se muestran los diagramas de confiabilidad para los umbrales de 1 y 25 mm. Se puede ver que en el caso de 1mm, hay una leve subestimación para las probabilidades inferiores al 30% y para las mayores probabilidades se ve una sobrestimación de las precipitaciones débiles para todos los ciclos. Mientras que para 10mm se ve un mejor desempeño en las probabilidades bajas y una sobrestimación más evidente en los ciclos de 06 y 18Z. Las curvas son irregulares debido a que la muestra de datos es pequeña.







Figura 6: Diagrama de confiabilidad para los umbrales 1 (izq.) y 10 mm (der.) para el ciclo 06Z. En la esquina superior izquierda se muestra el histograma de frecuencias de probabilidades pronosticadas para cada uno de los ciclos.

En la Figura 7 se pueden ver los diagramas ROC para el umbral de 1 y 10mm. Estos diagramas muestran la habilidad de los pronósticos en discriminar la ocurrencia o no de eventos dado cierto umbral de precipitación, por lo que indican la resolución del sistema de pronóstico (Murphy, 1993). La curva ROC muestra el POD (Probability of Detection) en función del POFD (probability of false detection) usando una serie de umbrales de probabilidad crecientes. El área bajo la curva (AUC, por sus siglas en inglés) se calcula mediante una aproximación trapezoidal. Un AUC igual a 0.5 indica que el sistema de pronóstico no tiene habilidad para discriminar la ocurrencia de eventos, si supera 0.8 se indica como un 'buen' pronóstico (Mullen and Buizza, 2002) mientras que si es 1 significa que es un pronóstico perfecto. No es sensible al bias sistemático por lo que no brinda información sobre la confiabilidad. Un sistema de pronóstico con bias sistemático igualmente puede presentar una buena resolución y por ende generar una buena curva ROC, significa que puede ser posible mejorar los errores mediante la calibración. Es un complemento del diagrama de confiabilidad ya que tiene más dependencia de las observaciones (dado que el evento se observó,¿cuál fue la probabilidad pronosticada?). A medida que el umbral de probabilidad disminuye, la probabilidad de detectar correctamente un evento (POD) aumenta pero la probabilidad de falsas detecciones (POFD) también aumenta. Se puede ver en la figura que para el umbral de 1mm hay una buena discriminación mientras que para 10mm bajan los valores de POD lo que indica que el sistema de pronóstico disminuye la cantidad de aciertos no logrando advertir la ocurrencia de dichos eventos.







Figura 7: ROC para los umbrales de 1 mm (izq.) y 10 mm (der.)(ciclo 06Z). El punto rojo indica el umbral de probabilidad óptimo para el cual se maximiza la diferencia entre el POD y el POFD. Se consideraron valores de PP24 superiores a 0.01mm.

3.2 Temperatura a 2 metros

En el análisis de la temperatura a 2m se exploró la relación entre el RMSE computado con el valor de la media del ensamble y el spread para todos los ciclos (Figura 8). Se puede ver por el signo del BIAS y la magnitud del RMSE que hay una sobrestimación de la T2m que es menor en el caso de la T2m calibrada para todos los plazos. La magnitud de la dispersión es levemente mayor a medida que el plazo aumenta para todos los ciclos aunque se ve un leve máximo relativo para las 18Z.

La magnitud de los resultados son similares a los que se obtienen con el SREF (NCEP <u>Short Range</u> <u>Ensemble</u> <u>Forecast</u> <u>System</u>) para el área continental de Estados Unidos (https://www.emc.ncep.noaa.gov/users/verification/regional/ensemble/ops/sfc_t2/).







Figura 8: Serie de RMSE de la media del ensamble (líneas continuas) y dispersión (líneas punteadas) para T2m (izq.) y T2m calibrada (der.) sobre todas las estaciones/fechas.



Figura 9: Serie de BIAS T2m (línea continua) y T2m calibrada (línea punteada) para el ciclo 06Z.





3.3 Temperatura mínima y máxima calibradas y sin calibrar

En cuanto a las temperaturas máximas y mínimas diarias, nuevamente se muestran los resultados para el ciclo de 06Z. A diferencia de los pronósticos del esquema determinístico, en este caso, dado que se cuentan con pronósticos hasta 48hs se puede hacer un análisis de los pronósticos para el DÍA+1 que corresponden a los valores pronosticados válidos entre 00 y 12Z para la Tmín y entre 12Z y 00Z para la Tmáx. Se puede ver en la figura 10 la distribución del RMSE de las temperaturas calibradas y sin calibrar tanto para WRF como para GFS del total de estaciones para todo el año. En primer lugar, se destaca la disminución de los errores debido a la calibración la cual tiene más impacto en las Tmáx. Asimismo, el WRF muestra un mejor desempeño que el GFS tanto para las temperaturas calibradas como sin calibrar.



Figura 10: Boxplots del RMSE de la Tmin (izq.) y Tmáx (der.) calibrada y sin calibrar para el WRF y GFS computados con los pronósticos válidos para el "día+1" entre 00-12z para la Tmin y entre 12-00z para la Tmáx (ciclo 06Z).

En particular, con el objetivo de explorar las distribuciones de los errores en las temperaturas calibradas, en la Figura 11 se pueden ver las distribuciones del RMSE para cada miembro del ensamble. La mediana de los errores de un conjunto de miembros ronda un valor superior. Estos miembros corresponden a la configuración de la representación de capa límite de MYJ la cual supone que no se produce mezcla con el aire por encima de la capa límite planetaria lo cual puede influir en las características termodinámicas de los niveles bajos.







Figura 11: Boxplots del RMSE de la Tmín calibrada (izq.) y Tmáx calibrada (der.) de los miembros del WRF válidos para el "día+1" (ciclo 06Z).

4. CONCLUSIONES

Este informe recopila los resultados de la verificación anual del año 2020 de los pronósticos de alta resolución del ensamble del modelo WRF-SMN. En particular, se analizaron las variables de precipitación acumulada en 24hs, temperatura a 2m calibradas y sin calibrar, temperatura diaria máxima y mínima calibradas y sin calibrar. En el análisis de las temperaturas 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.

En cuanto a la precipitación acumulada en 24 horas, la media del ensamble muestra resultados comparables a los hallados para el esquema determinístico mientras que analizando los miembros por separado se puede ver que logran una mejor correspondencia con la distribución de las observaciones que la media que refleja una distribución desplazada hacia las precipitaciones más débiles. Los eventos donde la precipitación es menor a 1mm quedan subestimados mientras que se sobrestiman para los umbrales mayores. El histograma de rangos reveló que el ensamble resultó subdispersivo dado que la distribución de los miembros queda ubicada sobre el centro de la muestra respecto de la variabilidad observada. Mientras que el diagrama de confiabilidad mostró que las





probabilidades más bajas son subestimadas y las superiores al 30% son sobrestimadas para el umbral de 1 y 10mm.

El RMSE y el BIAS de la temperatura a 2m muestran una sobrestimación para todos los ciclos que disminuye en el caso de la temperatura calibrada. Asimismo, la dispersión aumenta con el plazo con un leve máximo relativo a las 18Z.

En el caso de las temperaturas máximas y mínimas diarias del ensamble de WRF mostraron mejores desempeños que para el ensamble de GFS, asimismo que sus correspondientes variables calibradas mostraron errores menores que las no calibradas. También se exploró la distribución de errores de los miembros y se vio que aquellos con una configuración de capa límite de MYJ mostraron un error levemente mayor.

Se propone en trabajos futuros de verificación incorporar el uso de información proveniente de una base de datos de reanálisis de modelos independientes y estimaciones de precipitación satelital que puedan funcionar como proxy de las observaciones, con el objetivo de realizar una verificación en todo el dominio de pronóstico y con una resolución temporal adecuada a los pronósticos. Asimismo, con una mayor disponibilidad de datos se podría considerar una evaluación regionalizada con muestras climatológicamente homogéneas. Por otra parte, es de interés realizar una evaluación de otras variables como magnitud del viento en niveles bajos, o variables de humedad en cada uno de los miembros, con el fin de analizar el impacto de las distintas parametrizaciones de capa límite. Es necesario continuar con la exploración de medidas estadísticas adecuadas que logren describir el comportamiento de los pronósticos probabilísticos de modo que se refleje en forma objetiva el desempeño y habilidad del sistema. Asimismo, deben encontrarse las medidas adecuadas para que la calidad del ensamble refleje también su valor.

Agradecimientos: Al proyecto CyT Alerta con el cual se adquirió el HPC Huayra Muyu y a Melissa Patanella y Pablo Irurzun por facilitar un desarrollo en Python que genera el diagrama de desempeño.





REFERENCIAS

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

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.

Mullen, S. y Buizza, R., 2002: The Impact of Horizontal Resolution and Ensemble Size on Probabilistic Precipitation Forecasts by the ECMWF Ensemble Prediction System. AGU Spring Meeting Abstracts. 17. DOI: 10.1175/1520-0434(2002)017<0173:TIOHRA>2.0.CO;2.

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.

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





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_Olavarria	-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_Quiaca	-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_Chamical_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_lguazu	-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

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

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 del ensamble

Spread =
$$\sqrt{\frac{\sum_{i}^{M} (P_{i} - \underline{P})^{2}}{M-1}}$$

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

Es una medida de la variabilidad de los miembros dada por la distancia media de los miembros al valor de la media del ensamble.

Matriz de contingencia:

	Obse			
Pronóstico	SI	NO	Total	
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.

Estadísticos categóricos				
BIAS SCORE o Frequency BIAS (FBIAS) Indica la relación entre la frecuencia de eventos pronosticados y eventos observados. $0 \le FBIAS \ge \infty$; valor óptimo = 1. Valores negativos indican que el modelo subestima, valores positivos indican que el modelo sobreestima.	$FBIAS = \frac{A+F}{A+S}$			





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	$CSI = \frac{A}{A + S + F}$
POD (probability of detection) Indica la fracción de los eventos observados que se pronosticaron correctamente $0 \le POD \ge 1$; valor óptimo = 1	$POD = \frac{A}{A+S}$
FAR (false alarm ratio) Indica la fracción de los eventos pronosticados que no se observaron $0 \le FAR \ge 1$; valor óptimo = 0	$FAR = \frac{F}{A+F}$
POFD (probability of false detection) Indica la fracción de los eventos no observados que no fueron pronosticados $0 \le POFD \ge 1$; valor óptimo = 0	$POFD = \frac{F}{N+F}$

Estadísticos continuos

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.

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.
RMSE (root mean square error)
Indica una medida del valor medio de los errores
RMSE varía entre cero e infinito (0; ∞).
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (P_i - O_i)^2}{N}}$$



23



Distribución de probabilidad volumétrica de precipitación (PDF)

Las distribuciones de probabilidad del volumen de precipitación (Amitai etal., 2011), 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=10log(R/1mmdía-1); [R]=[mmdía-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 PPacum24h 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 Frequeny 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.



24



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.

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

donde *f* son los pronósticos probabilísticos 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).

Se puede descomponer en componentes llamadas, confiabilidad, resolución e incertidumbre (véase, por ejemplo, Wilks 1995, 261-263) como:

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



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 los 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 ascendiente (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.





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 (<u>rdelia@smn.gov.ar</u>), Luciano Vidal (<u>lvidal@smn.gov.ar</u>) o Martin Rugna (<u>mrugna@smn.gov.ar</u>) 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).

