

Verificación de los pronósticos probabilísticos de El Niño y La Niña costeros

Ken Takahashi
Instituto Geofísico del Perú

Una necesidad básica para la mejora de los pronósticos es una evaluación sistemática y objetiva de su desempeño (ver discusión en Takahashi, 2015). Una condición esencial, pero que para muchos no es obvia, es que los pronósticos deben ser verificables objetivamente. Por ejemplo, un pronóstico formulado como “se espera lluvia en la costa norte en los próximos días” no es verificable objetivamente, ya que no se especifica dónde, ni cuánto, ni cuándo. En contraste, decir “se espera que la lluvia exceda el percentil de 90 histórico en la ciudad de Piura en los próximos dos días” sí se puede verificar objetivamente. Si bien no todo pronóstico podrá ser igualmente verificable (por ej. para alguna región no existe un registro histórico de lluvias), es necesario que al menos parte de los pronósticos sean verificables objetivamente.

Para el pronóstico climático, que se realiza con varios meses de anticipación, existen dos limitantes para la verificación. En primer lugar, se realizan pocos pronósticos al año, por lo que desde un punto de vista simplemente estadístico se necesitaría esperar varios años para poder sacar conclusiones claras. En segundo lugar, los pronósticos estacionales son inciertos y se expresan como probabilidades de los diferentes escenarios posibles. Esto quiere decir que en teoría no es posible demostrar que un pronóstico es errado a menos de que este haya asignado 0% de probabilidad a lo que finalmente ocurrió.

Lo anterior, sin embargo, no es insalvable. Uno de los índices más populares de verificar un pronóstico probabilístico individual es el llamado “Brier score” (BS, ej. Wilks, 2006) que compara la probabilidad pronosticada p de un evento con la probabilidad o verificada, la cual es $o = 1$ si el evento ocurrió y $o = 0$ si no ocurrió (noten que una probabilidad de 1 corresponde a 100%). El BS para un pronóstico individual está dado por $BS = (p - o)^2$. Si se cuentan con varios pronósticos, entonces se puede calcular el valor promedio de BS sobre todos estos pronósticos. El pronóstico perfecto corresponde a $p = o$, o sea que $BS = 0$, mientras que el peor pronóstico posible es que se pronostique lo opuesto a lo que ocurrió con total certeza, es decir que $p = 1 - o$ por lo tanto $BS = 1$.

Empezando en el año 2015, el Comité ENFEN (ahora Comisión) empezó a producir pronósticos probabilísticos de la magnitud de El Niño/La Niña para los meses de verano (diciembre-marzo), en los que se consideran ocho posibles resultados para la región costera (La Niña fuerte, moderada y débil, neutral, El Niño débil, moderado, fuerte y extraordinario) y, similarmente, nueve para el Pacífico central (ENFEN, 2015). En esta situación no es posible utilizar el BS, ya que este considera solo un tipo de evento. Sin embargo, una generalización del BS es el “Ranked Probability Score” (RPS, ej. Tippet et al., 2017) permite trabajar con más de una categoría. En este caso, se considera la probabilidad acumulada pronosticadas (p) y observadas (o) de las categorías ordenadas de menor (La Niña fuerte) a mayor (El Niño extraordinario). Entonces, de esta manera, si consideramos que en el verano 2015-2016 se verificaron condiciones El Niño costero moderado (sexta categoría de ocho), entonces las probabilidades acumuladas correspondientes son $o = 0, 0, 0, 0, 0, 1, 1, 1$ (o sea, 1 a partir del sexto valor). Por otro lado, el pronóstico realizado por el ENFEN a finales de agosto de 2015 (ENFEN 2015) fue $p = 0.01, 0.01, 0.01, 0.01, 0.15, 0.45, 0.80, 1$. Entonces, el valor de RPS se calcula como la suma de los valores de $(p - o)^2$

sobre las categorías (en la práctica no se considera la última ya que siempre tiene $o = p = 1$). Es decir que para este caso, $RPS = (0-0)^2 + (0-0)^2 + (0-0)^2 + (0.05-0)^2 + (0.15-0)^2 + (0.45-1)^2 + (0.8-1)^2 = 0.37$. Al igual que con el BS, el pronóstico perfecto tendría $RPS = 0$ y el valor aumenta conforme el pronóstico es peor, pero el valor numérico de RPS en sí no es fácil de interpretar por sí solo. Sin embargo, sirve para comparar entre pronósticos para diferentes años o con diferentes métodos. En el segundo caso, si consideramos la frecuencia histórica observada (1982-2010) de las diferentes categorías podríamos hacer un pronóstico “climatológico” $p_{clim} = 0, 0, 0, 0.86, 0.93, 0.93, 0.97, 1$. Para el verano 2016-2017 (El Niño costero moderado) esto daría $RPS_{clim} = 1.62$. Como este valor es mayor que el correspondiente al del pronóstico del ENFEN, entonces podemos decir que este último fue mejor según el RPS.

Además, también se consideran los pronósticos probabilísticos obtenidos de las corridas de ocho modelos climáticos que participan en el proyecto *North American Multimodel Ensemble* (NMME; Kirtman et al., 2014). Para esto, primero a cada modelo se le restó el promedio climatológico (1982-2010) para las corridas inicializadas en cada mes del año y para todo lead (tiempo de anticipación) entre 0.5 y 8.5 meses, lo cual elimina los errores sistemáticos promedio (“mean bias correction”, ej. Tippet et al., 2017) se consideraron 10 corridas de cada modelo, dando un total de 80 corridas, y se determinó para cada corrida cuál es la categoría pronosticada siguiendo un procedimiento análogo al operacional (ENFEN 2012). Luego, se determinó la probabilidad para cada categoría básicamente como el número de corridas correspondientes dividido por el número total de corridas (con una corrección pequeña siguiendo a Tippet et al., 2017).

También se consideró una versión “calibrada” de cada modelo, para la que se aplicó una regresión lineal entre la anomalía observada y la pronosticada por cada modelo (promedio de *ensemble*) para cada mes inicial y lead en el periodo 1982-2010 para acercar los modelos a las observaciones.

En la Fig. 1 se muestran los valores de RPS para los pronósticos realizados para los veranos 2015-2016 y 2016-2017 entre los meses de agosto y setiembre. Podemos ver que los pronósticos realizados en 2015 por los modelos y el ENFEN fueron buenos comparados con el climatológico, mejorando ligeramente (RPS

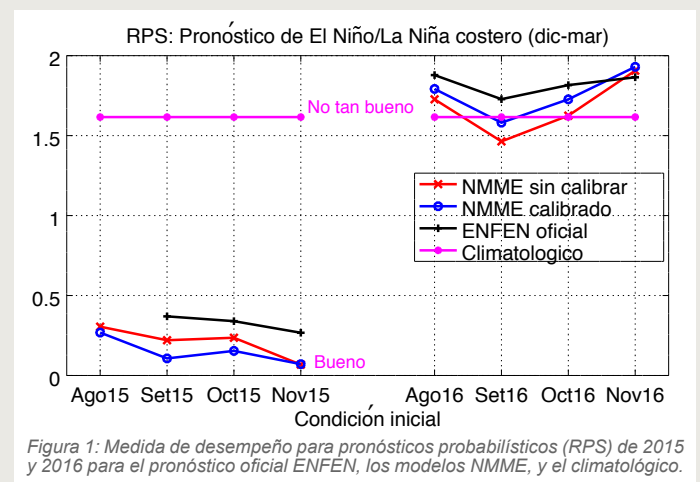


Figura 1: Medida de desempeño para pronósticos probabilísticos (RPS) de 2015 y 2016 para el pronóstico oficial ENFEN, los modelos NMME, y el climatológico.

Verificación de los pronósticos probabilísticos de El Niño y La Niña costeros

Ken Takahashi
Instituto Geofísico del Perú

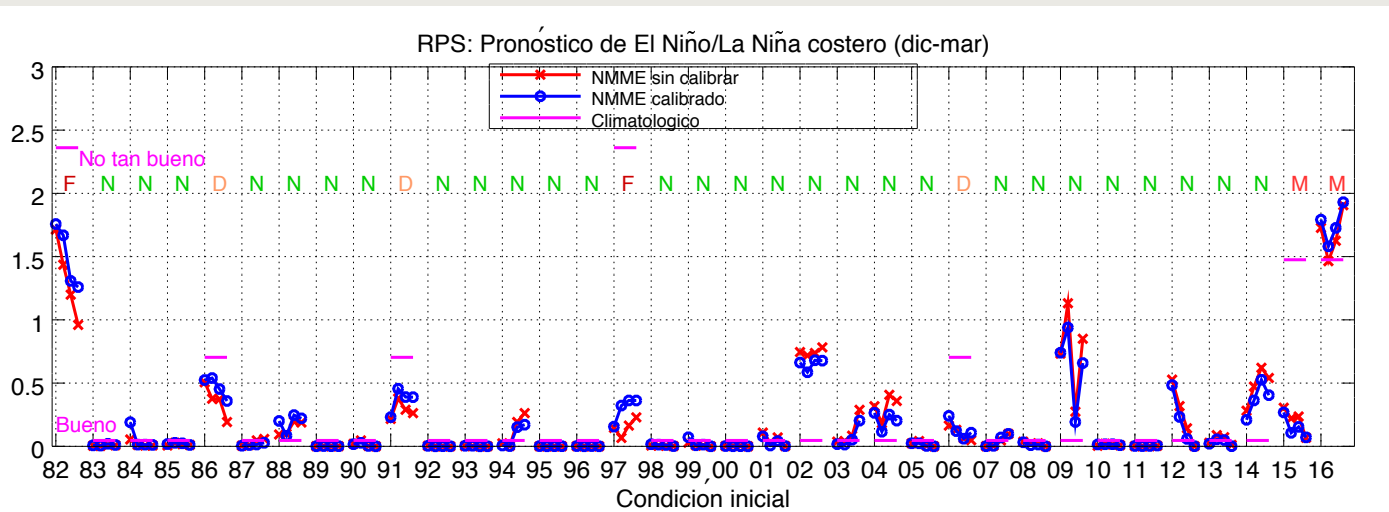


Figura 2: Medida de desempeño para pronósticos probabilísticos (RPS) entre 1982 y 2016 para el pronóstico de los modelos NMME y el climatológico. La magnitud de El Niño en la costa se indica con letras de colores.

disminuyendo) a medida que el año avanzó. Esto correspondió a condiciones El Niño de gran escala que aportaron mayor predictibilidad al sistema (L'Heureux et al., 2016). Los modelos tuvieron un mejor desempeño que el ENFEN ya que este último asignó mayor probabilidad a El Niño fuerte y extraordinario que los modelos. Por otro lado, en el año 2017 se presentó lo que ahora se conoce popularmente como “El Niño costero 2017” (ENFEN 2017), similar al ocurrido en 1925 (Takahashi y Martínez 2017), con mecanismos locales rápidos y menos predecibles. Ni los modelos climáticos ni el ENFEN previeron este evento y los pronósticos tuvieron desempeño similar al del pronóstico climatológico, aunque los de NMME tuvieron un desempeño ligeramente mejor (menor RPS), excepto en noviembre en el que ENFEN fue marginalmente mejor debido a que se asignó ligeramente mayor probabilidad de El Niño moderado que los otros. Vemos que, en general, la calibración de los modelos mejoró ligeramente su desempeño en 2015, pero no en el 2016. En promedio sobre estos ocho pronósticos, los modelos sin calibrar tuvieron mejor desempeño (RPS=0.94), seguido de los modelos calibrados (0.95), el ENFEN (1.18) y el climatológico (1.62).

Sobre el periodo de disponibilidad de los modelos NMME (1982-2016), el desempeño según el RPS promedio de los modelos calibrados y sin calibrar fueron casi iguales al climatológico (RPS=0.32 en todos los casos). Sin embargo, en la Fig. 2 vemos que este aparente buen desempeño se debe a que los pronósticos y las observaciones están dominados por condiciones neutras pero si limitamos el análisis a los años de eventos El Niño (1982, 1986, 1991, 1997, 2006 y 2015), exceptuando 2016, entonces vemos que los modelos tuvieron sustancialmente mejor desempeño (noten que no hubieron eventos La Niña de verano en todo este periodo). Por otro lado, en los años neutros en la costa pero cálidos en el Pacífico central (2002, 2009, 2014) los modelos pronosticaron mayor probabilidad de El Niño débil, y el pronóstico climatológico tuvo mejor desempeño.

Con respecto a los eventos El Niño de más impacto en la costa, los modelos tuvieron un desempeño relativamente bueno en el evento global de 1997, pero no tanto en el de 1982. Mientras tanto, El Niño costero de 2017 no fue pronosticado adecuadamente. Dado que cada evento tiene sus particularidades, no podemos confiar tan solo en los modelos y más bien es importante promover

la investigación científica para mejorar nuestro entendimiento y poder contar con mejores modelos conceptuales, así como numéricos.

Referencias

- ENFEN, 2012: Definición operacional de los eventos El Niño y La Niña y sus magnitudes en la costa del Perú. Nota Técnica ENFEN, 3 pp.
- ENFEN, 2015: Pronóstico probabilístico de la magnitud de El Niño costero en el verano 2015-2016, Nota Técnica ENFEN N°02-2015, 7 pp.
- ENFEN, 2017: El Niño Costero 2017. Informe técnico extraordinario ENFEN 001-2017, 31 pp.
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Peña Mendez, M., Becker, E., Peng, P., Tripp, P., Huang, J., Dewitt, D. G., Tippett, M. K., Barnston, A. G., Li, S., Rosati, A., Schubert, S. D., Rienecker, M., Suarez, M., Li, Z. E., Marshak, J., Lim, Y. K., Tribbia, J., Pegion, K., Merryfield, W. J., Denis, B., Wood, E. F., 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, doi: 10.1175/BAMS-D-12-00050.1.
- L'Heureux M, Takahashi K, Watkins AB, Barnston AG, Becker EJ, Di Liberto TE, Gamble F, Gottschalck J, Halpert MS, Huang B, Mosquera-Vásquez K, Wittenberg AT, 2016: Observing and predicting the 2015-16 El Niño. *Bull. Amer. Meteor. Soc.*, doi:10.1175/BAMS-D-16-0009.1
- Takahashi, K., 2015: La ciencia de la predicción. *Bol. Técnico El Niño IGP*, 2 (10), 4-8.
- Takahashi, K., Martínez, A. G. 2017: The very strong coastal El Niño in 1925 in the far-eastern Pacific. *Climate Dynamics*, doi: 10.1007/s00382-017-3702-1.
- Tippett, M. K., Barnston, A. G., Li, S., 2012: Performance of recent multimodel ENSO forecasts. *J. Appl. Met. Clim.*, doi:10.1175/JAMC-D-11-093.1.
- Tippett, M. K., Ranganathan, M., L'Heureux, M. L., Barnston, A. G., DelSole, T., 2017: Assessing probabilistic predictions of ENSO phase and intensity from the North American Multimodel Ensemble. *Climate Dynamics*, doi:10.1007/s00382-017-3603-3.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*, 2nd Edition, Academic Press, 627 pp.